

WO 93/03157(1)  
 (C07K15/00F2482)-(C07K15/00F30)-  
 C12N15/62-[M12K211:00]-[M12N211:00]-

PROPERTY ORGANIZATION  
 International Bureau



B 46

- 3 -

C12N15/62 UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :  
 C12N 15/62, 15/12, 9/10  
 C07K 13/00

A1

(11) International Publication Number: WO 93/03157  
 (43) International Publication Date: 18 February 1993 (18.02.93)

(21) International Application Number: PCT/US92/06187  
 (22) International Filing Date: 24 July 1992 (24.07.92)  
 (30) Priority data: 736,847 29 July 1991 (29.07.91) US  
 (71) Applicant: DANA FARBER CANCER INSTITUTE [US/US]; 44 Binney Street, Boston, MA 02115 (US).  
 (72) Inventors: BLANAR, Michael, A. ; 178 Locksley Avenue, Suite 1, San Francisco, CA 94122 (US). KAELIN, William, G. ; 12 Commonwealth Avenue, No. 301, Boston, MA 02116 (US).  
 (74) Agents: BROOK, David, E. et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

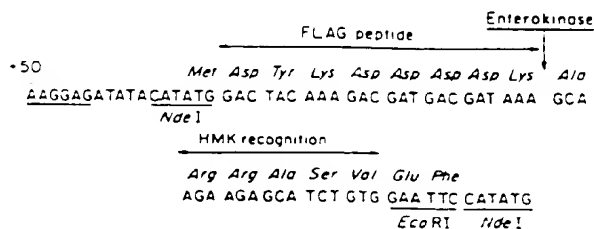
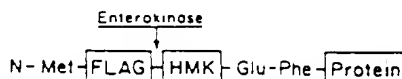
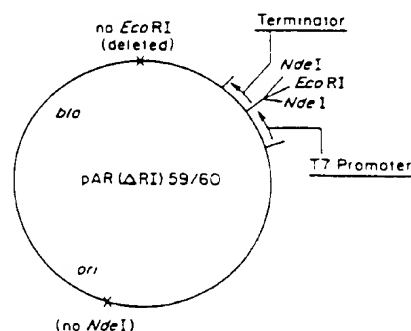
(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published  
 With international search report.

(54) Title: PLASMIDS FOR THE RAPID PREPARATION OF MODIFIED PROTEINS

(57) Abstract

The present invention relates to expression vectors comprising nucleic acid sequences which encode an affinity ligand (e.g., an enzyme, epitope) and a modification recognition sequence. The vectors further comprise at least one restriction site for the insertion of a nucleic acid sequence capable of encoding a selected polypeptide. On expression, the resulting construct codes for a fusion protein comprising an affinity ligand, the selected polypeptide and a modification recognition sequence. The fusion protein may be isolated by virtue of the affinity ligand and then modified. The expression vectors may further comprise a nucleotide sequence encoding a cleavable linker, such as a thrombin or factor Xa cleavage sequence. The invention further relates to expression vectors containing a nucleotide sequence encoding a gene for a selected polypeptide and capable of directing the expression of the selected polypeptide as a fusion protein. In addition, methods of producing a modified fusion protein are disclosed. Modified or labeled fusion proteins of the present invention are useful in a variety of therapeutic, diagnostic and research applications.



*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MN	Mongolia
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	GN	Guinea	NO	Norway
BG	Bulgaria	GR	Greece	NZ	New Zealand
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	PT	Portugal
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovak Republic
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CS	Czechoslovakia	LU	Luxembourg	SU	Soviet Union
CZ	Czech Republic	MC	Monaco	TD	Chad
DE	Germany	MG	Madagascar	TG	Togo
DK	Denmark	ML	Mali	UA	Ukraine
ES	Spain			US	United States of America

PLASMIDS FOR THE RAPID PREPARATION OF  
MODIFIED PROTEINS

Description

Background

05        Current methods for the production of labeled  
proteins include iodination, biotinylation, or the in  
vivo labeling of protein. In vivo labeling methods  
require large amounts of radioactivity are required,  
and the protein of interest must be separated from  
10 other labeled cellular proteins. The biotinylation  
and iodination procedures also require a purification  
step for each individual protein and, in some cases,  
reaction conditions which can cause inactivation of  
the protein. In addition, using these methods,  
15 modification of a protein may occur at a variety of  
sites, leading to distortions in structure or  
biological activity.

For example, iodination protocols relying on  
chloramine T or iodogen result in modifications at  
20 tyrosine residues and some histidine residues.  
Over-substitution and oxidation damage may result.  
Labeling procedures which use the Bolton-Hunter

-2-

reagent result in the modification of free amino groups on lysine residues. In some proteins this particular modification may also have a deleterious effect on structure or activity. Although the  
05 lactoperoxidase method for iodination employs gentler conditions, this method also leads to modification of tyrosine and histidine residues, with the potential for structural distortion and loss of activity.

Similarly, biotinylation protocols are  
10 frequently performed using a succinimide ester of biotin. The biotin is coupled to the protein through free amino groups, typically on lysine residues. Again, modification at one or more positions may alter structure and/or function of the protein. In  
15 addition, extensive dialysis is needed to remove uncoupled biotin, which may be deleterious to the protein.

#### Summary of the Invention

The present invention relates to expression  
20 vectors comprising nucleic acid sequences which encode an affinity ligand (e.g., an enzyme, epitope) and a modification recognition sequence. The vectors further comprise at least one restriction site for the insertion of a nucleic acid sequence capable of  
25 encoding a selected polypeptide in frame with the affinity ligand and modification sequence. On expression, the resulting construct codes for a fusion protein comprising an affinity ligand, the selected polypeptide and a modification recognition  
30 sequence.

-3-

The fusion protein may be isolated by virtue of the affinity ligand and then modified.

The expression vectors may further comprise a nucleotide sequence encoding a cleavable linker, such  
05 as a thrombin or factor Xa cleavage sequence. The sequence encoding the cleavable linker is located between the affinity ligand and the restriction site for insertion of the sequences encoding the selected polypeptide. In this location, the linker may be  
10 cleaved to release the protein of interest following modification. The invention further relates to expression vectors containing a nucleotide sequence encoding a gene for a selected polypeptide and capable of directing the expression of the selected  
15 polypeptide as a fusion protein.

The pGEX-2TK expression vector is one embodiment of the present invention. This vector encodes a protein comprising, from amino to carboxyl terminus, glutathione-S-transferase (GST) as an affinity  
20 ligand, the thrombin cleavage site as a cleavable linker, and a phosphorylation recognition site for the cAMP-dependent protein kinase as a modification recognition sequence. A multiple cloning site comprising three restriction sites is located  
25 downstream of the sequence encoding the phosphorylation site. Thus, a nucleic acid sequence encoding a selected polypeptide may be inserted into the vector using one or more of these sites. The selected polypeptide is expressed as a GTK-fusion  
30 protein (G, GST; T, thrombin; K, kinase).

-4-

In addition, methods of producing a modified fusion protein are disclosed. Upon expression in a suitable host cell, the protein of interest is produced as a fusion protein. The fusion protein can  
05 be captured on a suitable affinity matrix by virtue of an affinity ligand, which interacts reversibly with the matrix. Modification of the fusion protein can be carried out while the protein is attached to the matrix. Subsequently, the modified fusion  
10 protein may be isolated for use by releasing the fusion protein from the affinity matrix with a suitable agent. In the case where a cleavable linker is present, the fusion protein may be cleaved in vitro to free the modified polypeptide portion, and  
15 the affinity ligand portion or any uncleaved product can be removed by adsorption on the appropriate affinity matrix. Alternatively, the modified polypeptide portion can be released from the affinity ligand portion by cleaving the modified fusion  
20 protein at the cleavable linker while still bound to the column.

Modified proteins of the present invention are useful in a variety of applications. Labeled (e.g., radiolabeled) proteins may be used as molecular  
25 probes. For example, antibodies can be labeled for therapeutic, diagnostic (e.g., imaging) or research purposes. Proteins may be labeled and used as reagents to quantitate or identify an interacting protein, such as a receptor.

-5-

Brief Description of the Drawings

Figure 1 shows a portion of the nucleotide sequence around the kinase recognition site of expression vector pGEX-2TK. The portion of the  
05 pGEX-2TK sequence introduced by the synthetic duplex is indicated by italics. The sequences which encode the linker cleavable by thrombin (Leu-Val-Pro-Arg-Gly-Ser), the downstream kinase recognition site (Arg-Arg-Ala-Ser-Val) and multiple cloning site with  
10 BamHI, SmaI and EcoRI sites, are shown. The arrow indicates the point of thrombin cleavage.

Figure 2 is an illustration of the structure of expression vector pAR( $\Delta$ RI)59/60. The general structure of the plasmid is shown at top. bla,  
15 indicates the  $\beta$ -lactamase gene which confers ampicillin resistance; ori, indicates the origin of replication. In the center panel, the general structure of FEK-fusion proteins is illustrated. The FLAG peptide portion, comprising the FLAG epitope and enterokinase cleavable linker, is indicated. "HMK"  
20 indicates the location of the phosphorylation site. "Protein" indicates the location of the selected polypeptide. The lower panel shows a more detailed illustration of the N-terminal region of the vector.  
25 The peptide sequence shown (Met-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-Ala-Arg-Arg-Ala-Ser-Val-Glu-Phe-) in the detail is a contiguous sequence. The extent of the FLAG peptide, the enterokinase cleavage site, and HMK recognition (phosphorylation site) are indicated.

30 Figure 3 is a bar graph showing the effect on <sup>32</sup>P incorporation in vitro of an HMK sequence on

-6-

different fusion proteins, comprising portions of the retinoblastoma susceptibility gene product, RB. Proteins were phosphorylated in vitro using CAMP-dependent protein kinase. The hatched bars  
05 indicate that the fusion protein (GST-RB379-792 or GST-RB379-792;pm706) was expressed from pGEX-2TK and contained an HMK sequence, while the solid bars indicate that the fusion protein (GST-RB379-792 or GST-RB379-792;pm706) was expressed from pGEX-2T and  
10 did not contain an HMK sequence.

Unless indicated otherwise, the orientation of particular amino acid sequences is such that the amino end is on the left and the carboxyl end is on the right.

#### 15 Detailed Description of the Invention

The present invention relates to expression vectors comprising nucleic acid sequences which encode an affinity ligand and a modification  
recognition sequence. The vectors further comprise  
20 at least one restriction site for the insertion of a nucleic acid sequence capable of encoding a selected polypeptide. The expression vectors may further comprise a nucleotide sequence encoding a cleavable linker sequence. These vectors are referred to as  
25 parent vectors.

The present invention further relates to expression vectors which are derived from the parent vectors described above, by the insertion of a nucleic acid sequence capable of encoding a selected  
30 polypeptide (e.g., a natural or synthetic cDNA or

-7-

genomic DNA which encodes the selected polypeptide and is capable of being expressed in the appropriate host cell) into the parent vector. The parent vector can be cleaved at one or more restriction  
05 sites in order to insert sequences encoding a selected polypeptide using known techniques. Linkers or other sequences can be used to facilitate insertion. Insertion at an appropriate restriction site or site in the parent vector results in the in  
10 frame fusion of the sequences of the selected polypeptide with the amino acid sequences for an affinity ligand, modification recognition sequence and, if present, the optional cleavable linker encoded by the parent vector. On expression, the  
15 resulting fusion gene encoded by the vector codes for a fusion protein comprising an affinity ligand, a modification recognition sequence, selected polypeptide, and optionally, a cleavable linker. The location of the restriction site or sites selected  
20 for insertion in the vector will determine the location of the selected polypeptide relative to the affinity ligand, modification recognition sequence and optional cleavable linker in the encoded fusion protein. The selected polypeptide element of the  
25 fusion protein comprises at least one peptide, polypeptide or protein of interest.

The fusion protein comprising an affinity ligand, a modification recognition sequence, an optional cleavable linker and a selected polypeptide  
30 forms a contiguous polypeptide chain. The order of these components in the fusion gene and protein can vary. The location of these components or additional

-8-

components in the fusion gene and protein is determined by the order in which the nucleic acid sequences encoding the components are linked in the vector or the parent vector. For example, fusion  
05 proteins of the following structure can be made, where A indicates the affinity ligand, M indicates the modification recognition sequence, C indicates the optional cleavable linker, and X indicates the polypeptide of interest, and the N-terminal portion  
10 of the fusion is on the left:

A-(C)-M-X  
A-M-(C)-X  
X-(C)-M-A  
X-M-(C)-A  
15 M-A-(C)-X  
M-(C)-A-X

Other permutations of the components are possible. However, the cleavable linker is preferably located between the affinity ligand and selected polypeptide  
20 portions of the fusion protein. In addition, introduction of a selected polypeptide in frame is simplified by a terminal location for the polypeptide (X). This location can also minimize the effect of fusion on the biological activity (e.g., binding  
25 activity, antigenicity, enzymatic activity) of the selected polypeptide.

One or more components can be duplicated (e.g., X-M-C-A-X). For example, multiple modification sites can be incorporated to increase the intensity of

-9-

labeling. These sites may be contiguous or noncontiguous in the encoded fusion protein. Furthermore, additional sequences (S) can be present in the fusion protein. For instance, a sequence  
05 encoding a signal peptide or leader peptide may be incorporated into the vector, and the signal peptide will be produced as part of the fusion protein. In the case of a signal peptide, the additional sequence is preferably incorporated at the N-terminus (e.g.,  
10 S-X-M-C-A). However, for other sequences, an internal or C-terminal location in the fusion protein may be desired.

Note that on expression in a suitable host cell, the parent vector may also produce the affinity  
15 ligand, modification sequence, and optional cleavable linker sequences as a fusion protein. This product may contain additional sequences encoded by the vector, depending on the location of a promoter or termination signals relative to the coding sequences  
20 for these elements. In some cases, restriction sites for insertion of sequences encoding a selected polypeptide may disrupt the reading frame of components encoded by the parent vector. Insertion of sequences encoding a selected polypeptide will  
25 restore the reading frame.

The expression vectors of the present invention can be designed for use in a variety of host cells, including bacterial host cells such as E. coli and eukaryotic host cells (e.g., yeast cells and  
30 mammalian cells). Thus, fusion proteins can be glycosylated. Like other expression vectors, in the

-10-

expression vectors of the present invention, a promoter is provided for expression of the fusion protein in a suitable host cell. Suitable promoters can be constitutive or inducible. In the vectors, 05 the promoter is operably linked to nucleic acid sequences encoding the fusion protein and is capable of directing the expression of the corresponding polypeptide. A variety of suitable promoters for procaryotic (e.g., lac promoter, tac promoter) and 10 eukaryotic hosts (e.g., yeast alcohol dehydrogenase (ADH1), SV40) are available.

In addition, the expression vectors typically comprise a selectable marker for selection of host cells carrying the plasmid and an origin of 15 replication, in the case of a replicable expression vector. Genes encoding products which confer antibiotic resistance are common selectable markers and may be used in prokaryotic (e.g.,  $\beta$ -lactamase for ampicillin resistance, tetracycline resistance) and 20 eukaryotic cells (e.g., G418). Genes encoding the gene product of auxotrophic markers (e.g., LEU2 and URA3) are commonly used as selectable markers in yeast. Use of viral or phage vectors, and vectors which are capable of integrating into the genome of 25 the host cell, such as retroviral vectors, are also contemplated. The present invention also relates to cells carrying these types of expression vectors (e.g., transformed cells).

The incorporation of one or more modification 30 recognition sequences into the fusion protein allows the production of modified fusion proteins. As shown in the Examples, incorporation of a detectable label

-11-

(e.g., radioactive, fluorescent) at the modification site provides a convenient way to label the fusion protein. Thus, the invention further relates to the fusion proteins and modified (e.g., labeled) fusion proteins produced from expression vectors of the present invention. It is also possible to produce fusion proteins of the present invention using methods of peptide synthesis, or in vitro transcription/translation procedures. Modification of synthetic fusion proteins is also possible.

The affinity ligand present in the fusion proteins of the present invention is a polypeptide encoding an affinity ligand. The affinity ligand (i.e., an affinity ligand or portion thereof) is a member of a specific binding pair and is capable of binding to a specific binding partner. Preferably, the binding of the affinity ligand to its specific binding partner is reversible, allowing recovery of fusion proteins comprising the affinity ligand where desired. Examples of affinity ligands include, but are not necessarily limited to, antibodies or portions thereof, antigens (e.g., influenza hemagglutinin) or epitopes (e.g., FLAG epitope), enzymes (e.g., glutathione-S-transferase (GST),  $\beta$ -galactosidase (lacZ), the trpE product), hormones, growth factors or other proteins capable binding to a specific binding partner (e.g., maltose binding protein, histidine hexamer (a heavy metal binding element), and protein A). A specific affinity ligand is an affinity ligand comprising that protein or peptide or a portion thereof. Thus, a glutathione-S-transferase affinity ligand is an

-12-

affinity ligand comprising glutathione-S-transferase or a portion thereof, which is capable of binding the specific binding partner (e.g., a substrate). The affinity ligand portion of the fusion protein is  
05 useful for a variety of purposes, such as providing a moiety for attachment to a support or to facilitate purification or identification.

For example, the fusion proteins can be captured on an appropriate affinity matrix. An  
10 affinity matrix is a solid support to which is attached (preferably covalently) a specific binding partner. Fusion proteins of the present invention, comprising an affinity ligand, can bind to a specific binding partner via the affinity ligand part of the  
15 fusion protein. In the case of an affinity ligand which is an antigen, an antibody or portion thereof can be used as a specific binding partner in an affinity matrix. Alternatively, an antigen or hapten may be incorporated into an affinity matrix, and used  
20 to capture fusion proteins comprising an antibody as an affinity ligand. A number of affinity ligand/affinity matrix pairs are available. Glutathione-S-transferase fusion proteins may be captured on immobilized glutathione as an affinity  
25 matrix, with glutathione as the specific binding partner. For example, glutathione sepharose (Pharmacia) or glutathione agarose beads (Sigma Chemical Corp.) can be used. Fusion proteins comprising a protein A, maltose binding protein, FLAG peptide, or a histidine hexamer affinity ligand can  
30 be captured on IgG Sepharose 6FF (Pharmacia), anylose resin (New England Biolabs), anti-FLAG M1 antibody or

-13-

anti-FLAG M2 antibody affinity resins (International Biotechnologies, Inc.), or a  $\text{Ni}^{2+}$  affinity resin (NTA resin, Qiagen), respectively. In the foregoing, IgG, amylose, anti-FLAG antibodies, or a heavy metal are, respectively, the specific binding partners.

When desired, a fusion protein or modified (e.g. labeled) fusion protein bound to an affinity matrix can be released by contacting the affinity matrix with bound fusion protein thereto with a suitable elution buffer comprising one or more release components. The release component or components can be molecules which compete with the fusion protein for binding to the affinity matrix (e.g., hapten, free peptide epitopes, substrate or substrate analogs), or which can disrupt binding of the affinity ligand to the specific binding partner. For example, an elution buffer (a buffered solution) suitable for releasing fusion proteins comprising a glutathione-S-transferase affinity ligand can be formulated comprising reduced glutathione as a release component. In one embodiment, an elution buffer comprising reduced glutathione, Tris and NaCl is used.

Alternatively, the elution buffer may comprise a buffered solution which lacks a specific component required for binding. For example, a fusion protein with an ompA signal sequence followed by a FLAG affinity ligand at the amino terminus can be expressed in E. coli. Specific removal of the ompA signal sequence upon secretion into the periplasmic space results in a fusion protein with an N-terminally located FLAG affinity ligand, which is

-14-

capable of binding to the anti-FLAG M1 antibody in the presence of calcium. This type of fusion protein can be eluted from an anti-FLAG M1 antibody affinity matrix using an elution buffer which lacks calcium,  
05 a specific component required for binding.

The expression vectors of the present invention can further comprise a sequence encoding a cleavable linker. A nucleotide sequence capable of encoding such a cleavage site can be incorporated into an  
10 expression vector using known techniques (e.g., recombinant DNA and/or mutagenesis). A cleavable linker is a peptide or polypeptide capable of being cleaved by a site specific protease. For example, a thrombin cleavage (e.g., Leu-Val-Pro-Arg-Gly-Ser),  
15 factor Xa cleavage site (e.g., Ile-Glu-Gly-Arg) or enterokinase cleavage site (e.g., Asp-Asp-Asp-Asp-Lys) may be used as a cleavable linker. The appropriate protease (thrombin, factor Xa and enterokinase, respectively) can be used to cleave a fusion  
20 protein containing a corresponding cleavable linker. Cleavage by a protease (proteolysis) can occur within a cleavable linker or at the border of the linker and another component of a fusion protein.

The product of the cleavage reaction of a  
25 selected fusion protein comprising the majority of the affinity ligand is referred to as the affinity ligand portion, and the product of the cleavage reaction comprising the selected polypeptide is referred to as the selected polypeptide portion of  
30 the fusion protein. These portions may comprise other parts of the encoded fusion protein (i.e., the

-15-

fusion protein encoded by the expression vector), and such are also fusion proteins. For example, the selected polypeptide portion can include the modification recognition sequence. If so, and if the  
05 fusion protein has been modified (e.g., labeled), then the selected polypeptide portion is referred to as the labeled selected polypeptide portion. Thus, the term fusion protein refers to a protein, polypeptide, or glycoprotein comprising at least two  
10 of the components selected from the group consisting of an affinity ligand, modification sequence, cleavable linker sequence, selected polypeptide or additional sequence.

Preferably, a cleavable linker is selected which  
15 does not cleave elsewhere within the fusion protein (e.g., in the selected polypeptide, affinity ligand or modification site). In addition, the sequences encoding the cleavable linker are preferably located between those sequences encoding the affinity ligand  
20 and selected polypeptide. In this location, the encoded fusion protein can be cleaved to separate the affinity ligand portion of the fusion protein from the selected polypeptide portion where desired. Similarly, where recovery of a labeled selected  
25 polypeptide portion is desired, the sequences encoding the cleavable linker will be on located in the vector either upstream or downstream of the sequences encoding both the modification recognition sequence and the selected polypeptide (e.g., as in  
30 A-C-M-X or X-M-C-A fusions).

Overlap of the sequences of portions of the fusion protein may occur provided the respective portions retain function. For example, in one

-16-

embodiment of the present invention, pAR( $\Delta$ R1)59/60, discussed in more detail below, the affinity ligand (FLAG epitope) sequence and cleavable linker sequence (enterokinase site) overlap by one amino acid. The  
05 sequence of the FLAG epitope from N- to C-terminus is (Asp-Tyr-Lys-Asp), while the enterokinase cleavage site in this embodiment is (Asp-Asp-Asp-Asp-Lys). Within pAR( $\Delta$ R1)59/60, these sequences overlap to give the FLAG peptide, an octapeptide of the sequence  
10 Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys, which retains the functions of the affinity ligand (e.g., binding to a specific binding partner) and cleavable linker (e.g., cleavage by enterokinase). Similarly, other components of the fusion protein can overlap,  
15 providing their function is preserved.

A modification recognition sequence is also provided. This sequence, incorporated into the fusion protein, directs modification of the fusion protein. Modification recognition sequences can be  
20 incorporated into a fusion protein comprising a selected polypeptide which either is naturally modified or is not naturally modified. A modification recognition sequence for phosphorylation (i.e., a phosphorylation site) can be introduced into  
25 a vector and expressed as part of a fusion protein, for example. In the presence of a suitable protein kinase, the fusion protein comprising the phosphorylation (kinase) site will be phosphorylated. Other types of modifications directed by the presence  
30 of a peptide or polypeptide sequence are envisioned in the present invention. For example, glycosylation

-17-

reactions can be directed by a short peptide sequence. Similarly, fatty acylation of some proteins is directed by a short peptide sequence (Cys-Ala-Ala-X). The identification of additional  
05 recognition sites and the identification and purification of the corresponding modification enzyme or enzyme complex will provide alternative protocols for modification of the fusion proteins.

Modification of the fusion proteins is carried  
10 out using a modification enzyme or enzyme complex or other suitable means. In the case of a phosphorylation reaction, a suitable phosphorylation method is used. Typically, the phosphorylation of proteins is carried out by a protein kinase. Many  
15 such kinases have been described and have utility in the present invention (see e.g., Kemp, B.E., et al., J. Biol. Chem., 252: 4888-4894 (1977); Edelman, A.M. et al., Ann. Rev. Biochem. 56: 567-613 (1987); Glass, D.B. and E.G. Krebs, Ann. Rev. Pharmacol. Toxicol.  
20 20: 363-388 (1980); Hunter, T. and J.A. Cooper, Ann. Rev. Biochem. 54: 897-930 (1985); Hunter, T., Cell 50: 823-829 (1987)). Known protein kinases catalyze the transfer of the  $\gamma$ -phosphate group of ATP to the  
25 hydroxyl groups of serine and/or threonine residues on specific substrates or alternatively, catalyze the phosphorylation of tyrosine residues. Use of protein kinases of both specificities is contemplated in the present invention. Serine/threonine kinases include cyclic AMP (cAMP) dependent and cyclic GMP (cGMP)  
30 dependent protein kinases, and cyclic nucleotide-independent protein

-18-

kinases. A variety of serine/ threonine kinases have been identified including glycogen synthase kinase, phosphorylase kinase, casein kinase I casein kinase II, pyruvate dehydrogenase kinase, protein kinase C  
05 and myosin light chain kinase.

Amino acid sequences present in natural substrates and artificial peptide substrates which are sufficient for activity as a protein kinase substrate have been identified (see e.g., Edelman,  
10 A.M. et al., Ann. Rev. Biochem. 56: 567-613 (1987); Glass, D.B. and E.G. Krebs, Ann. Rev. Pharmacol. Toxicol. 20: 363-388 (1980); Hunter, T. and J.A. Cooper, Ann. Rev. Biochem. 54: 897-930 (1985)). These amino acid sequences, in addition to those described  
15 below, and related amino acid sequences which can be specifically phosphorylated, can be used as modification recognition sequences in vectors of the present invention. A corresponding protein kinase, capable of phosphorylating a polypeptide comprising  
20 the selected kinase recognition sequence, can be used to phosphorylate fusion proteins of the present invention. For example, the cAMP-dependent protein kinase (e.g., the catalytic subunit of the cAMP-dependent protein kinase from bovine heart  
25 muscle) can recognize the consensus amino acid sequence Arg-Arg-Xaa-Ser-Xaa, where Xaa is an amino acid, in a variety of substrates, resulting in phosphorylation of serine. As shown in the Examples, expression vectors of the present invention (e.g.,  
30 PGEX-2TK derivatives, or pAR( $\Delta$ RI)59/60 derivatives), incorporating a sequence which encodes a version of

-19-

the cAMP-dependent protein kinase recognition sequence (Arg-Arg-Ala-Ser-Val), can direct the production of fusion proteins capable of being phosphorylated by cAMP-dependent protein kinase from  
05 bovine heart muscle. The Arg-Arg-Ala-Ser-Val sequence is also referred to as the HMK site (HMK, heart muscle kinase). Other specific amino acid sequences, such as Arg-Arg-Ala-Ser-Leu (Li et al., Proc. Natl. Sci. USA 86: 558-562 (1989) and the  
10 peptide Arg-Thr-Lys-Arg-Ser-Gly-Ser-Val, can be recognized and phosphorylated by this kinase. The phosphorylation sites can function at a terminal or internal location within a fusion protein.

cGMP protein kinases have a substrate  
15 specificity that is similar, but not identical, to that of the cAMP-dependent protein kinases. Comparative analysis of substrates has been made (Edelman, A.M. et al., Ann. Rev. Biochem. 56: 567-613, (1987); Glass, D.B. and E.G. Krebs, Ann.  
20 Rev. Pharmacol. Toxicol. 20: 363-388 (1980); Glass, D.B. and E.G. Krebs, J. Biol. Chem. 254: 9728-9738 (1979)). The substrate specificities of the two types of casein kinases has also been studied, and the activity of peptide substrates has been compared  
25 (Marin, O. et al., Eur. J. Biochem. 160: 239-244 (1986); Sommercorn, J. and E.G. Krebs, J. Biol. Chem. 262: 3839-3843 (1987); Kuenzel, E.A. et al., J. Biol. Chem. 262: 9136-9140 (1987)). For example, casein  
30 kinase II was shown to phosphorylate the synthetic peptide Ser-Glu-Glu-Glu-Glu. Additional peptide substrates were phosphorylated by casein kinase II (e.g., in decreasing order of activity,

-20-

Arg-Arg-Arg-Asp-Asp-Asp-Ser-Asp-Asp-Asp; Arg-Arg-Arg-Glu-Glu-Glu-Ser-Glu-Glu-Glu; Arg-Arg-Arg-Glu-Glu-Glu-Thr-Glu-Glu-Glu). The substrate specificity of tyrosine kinases has also been studied (Hunter, T. and J.A. Cooper, Ann. Rev. Biochem. 54: 897-930 (1985)). Candidate phosphorylation sites (recognition sequences) can be incorporated into synthetic peptides or into fusion proteins and assayed for activity as protein kinase substrates, using techniques similar to those described previously. Because corresponding phosphatases exist, enzymatic removal of phosphate groups is also possible.

The ability to modify the fusion proteins provides a convenient method of specifically labeling the fusion proteins with detectable radioactive or non-radioactive labels to produce a labeled fusion protein (e.g., a [ $^{32}\text{P}$ ]-labeled fusion protein). Cleavage of fusion proteins can produce a selected polypeptide portion comprising a modification sequence or an affinity ligand portion comprising a modification sequence, depending on the location of the cleavable linker in relation to these components. Thus, cleavage following labeling can produce a labeled selected polypeptide portion or a labeled affinity ligand portion, each of which is itself a fusion protein. Because the modification recognition sequence directs modification to a specific site in the protein, there is a high degree of control over the location and extent of modification. As shown in

-21-

the Examples, fusion proteins expressed and modified by the introduction of a detectable label retain both structure and function.

The detectable label (e.g., radioactive, 05 fluorescent or chemiluminescent) selected will determine by the nature of modification, the modification enzyme, and the intended use. The label will be incorporated into a moiety which is transferred to the fusion protein substrate. For 10 phosphorylation reactions, [ $\gamma$ -labeled]ATP is used as phosphate donor (i.e., modification donor). As protein kinases transfer the  $\gamma$ -phosphate onto the substrate during the phosphorylation reaction, the label will be incorporated into the moiety 15 transferred by the protein kinase. For example, to incorporate a radioactive phosphate label such as  $^{31}\text{P}$ ,  $^{32}\text{P}$  or  $^{33}\text{P}$ , phosphate donors such as [ $\gamma$ - $^{31}\text{P}$ ]ATP, [ $\gamma$ - $^{32}\text{P}$ ]ATP, or [ $\gamma$ - $^{33}\text{P}$ ]ATP can be used. Alternatively, isotopes of sulfur, such as  $^{35}\text{S}$  or  $^{38}\text{S}$  isotopes, can 20 be incorporated as the label in the [ $\gamma$ -labeled]ATP, as for example in  $^{35}\text{S}$ -labeled adenosine 5' [ $\gamma$ -thio]triphosphate. The term "phosphorylation" also refers to modification with such thiophosphate analogs or other  $\gamma$ -labeled analogs of ATP. Other 25 considerations such as specific activity, half-life, type of particle emitted and energy of radiation will influence selection of an appropriate radioactive label. In addition, a fluorescent moiety or chemiluminescent moiety incorporated into the 30  $\gamma$ -phosphate. Detection methods for such labels are well known in the art.

-22-

In other types of modifications (e.g., glycosylation, fatty acylation), a radioactive isotope, chemiluminescent or fluorescent dye, or other nonradioactive label, could be incorporated into the  
05 modification donor so that it is transferred to the fusion protein during the modification reaction. For example, a biotin adduct of a modification donor could be linked to a fusion protein. In the present method, the site of addition of the biotin label is  
10 controlled by the location of the modification recognition sequence.

#### Specific Embodiments

In one embodiment of the present invention, glutathione-S-transferase (GST) from the parasitic  
15 helminth Schistosoma japonica (Mr=26,000) is selected as the affinity ligand. However, GST genes from other sources, such as other bacteria or mammalian organisms, can be used. In addition, a portion of a GST gene encoding a portion of GST capable of binding  
20 a specific binding partner, such as the substrate glutathione, can be used. In particular, expression vectors capable of expressing S. japonicum GST-fusion proteins (i.e., a fusion protein comprising a GST affinity ligand) were constructed. The parent vector  
25 is named pGEX-2TK, and vectors derived from pGEX-2TK by the insertion of a nucleotide sequence encoding a selected polypeptide, are referred to with the prefix pGTK-. The construction of pGEX-2TK and pGTK-plasmids is described in Examples 1 and 2 and in  
30 Figure 1.

-23-

pGEX-2TK is a derivative of pGEX-2T, the latter which is described by D.B. Smith in EP 0,293,249, published November 30, 1988, PCT/AU88/00164, published December 1, 1988, and New Zealand Patent No. 224,663, issued 05 November 27, 1990, and by D.B. Smith and K.S. Johnson in Gene 67: 31-40 (1988). The teachings of EP 0,293,249, PCT/AU88/00164, New Zealand Patent No. 224,663, and Smith and K.S. Johnson (Gene 67: 31-40 (1988)) are herein incorporated by reference.

10 pGEX-2TK encodes a fusion protein having GST as an affinity ligand at the amino terminus, followed by a thrombin cleavage site (cleavable linker). In addition, a modification recognition sequence for a protein kinase was incorporated downstream of the 15 cleavable linker. In particular, a sequence of the structure Arg-Arg-Xaa-Ser-Xaa, where Xaa is an amino acid, was selected (Arg-Arg-Ala-Ser-Val). This sequence is a phosphorylation site recognized by cAMP-dependent protein kinases such as the catalytic 20 domain of the cAMP-dependent protein kinase from bovine heart muscle. In pGEX-2TK, the sequence which encodes the thrombin cleavage site is followed by several restriction sites for the insertion of sequences encoding a selected polypeptide. pGEX-2TK 25 retains the inducible tac promoter for expression, three in frame stop codons, a selectable marker (ampicillin resistance), origin of replication, and lacIq gene present in pGEX-2T.

30 When a sequence which encodes a selected polypeptide is inserted in frame into the pGEX-2TK expression vector downstream of the phosphorylation

-24-

site, a GTK-fusion protein comprising the selected polypeptide can be produced. The GTK-fusion protein (where, from N- to C-terminus, G is a GST affinity ligand, T is a thrombin cleavage site, and K is a phosphorylation site for a protein kinase) is encoded by a pGTK-vector, can be produced upon expression in a suitable E. coli host. These fusion proteins can be captured on immobilized glutathione, and labeled by phosphorylation with CAMP-dependent protein kinase, using [ $\gamma$ -<sup>32</sup>P]ATP as a modification donor. As shown in the Examples, several different cDNAs were expressed from pGEX-2TK, and labeled as described. The studies described in Examples 1-4 indicate that a selected polypeptide present in a GTK-fusion protein, encoded by a pGTK-type vector, such as pGTK-RB (379-792), can retain both structure and function. Furthermore, the labeling (phosphorylation) procedure was efficient and did not alter the structure or function of the selected polypeptides.

In another embodiment of the present invention, the FLAG epitope is selected as the affinity ligand. Vector pAR( $\Delta$ RI)59/60, shown in Figure 2, is an example of this type of expression vector. Vector pAR( $\Delta$ RI)59/60 is a derivative of bacterial expression plasmid pET3a (pAR3040) described by Studier et al., Meth. Enzymol. 185:60-89 (1990), the teachings of which are herein incorporated by reference. The construction of pAR( $\Delta$ RI)59/60 is described in detail in Example 5. In this construct, the FLAG peptide is located adjacent to the N-terminal initiator

-35-

methionine (Figure 2). The FLAG octapeptide (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys) comprises the FLAG epitope (Asp-Tyr-Lys-Asp) affinity ligand and an overlapping enterokinase cleavable linker (Asp-Asp-Asp-Asp-Lys). Cleavage with enterokinase typically occurs precisely after the terminal lysine (underlined above) of the FLAG octapeptide. The FLAG octapeptide is followed by a kinase recognition site. In particular, in pAR( $\Delta$ RI)59/60, the site is Arg-Arg-Ala-Ser-Val (HMK), recognized by cAMP-dependent protein kinase. A unique EcoRI restriction site located downstream of the sequence encoding the phosphorylation site can be used for the insertion of a sequence encoding a selected polypeptide. In plasmids derived from pAR( $\Delta$ RI)59/60, referred to with the prefix pFEK- F for FLAG, E for enterokinase, and K for the phosphorylation recognition site), in which a sequence encoding a selected polypeptide has been inserted, the encoded fusion protein, referred to herein as a FEK-fusion protein, has the following general structure for N- to C-terminus: Met-[FLAG/enterokinase cleavable linker]-phosphorylation site-selected polypeptide. An alanine residue is located between the FLAG peptide and HMK site (see Figure 2). In addition, the EcoRI site encodes a Glu-Phe dipeptide. Depending on the strategy for insertion of a sequence encoding a selected polypeptide, additional residues (e.g., encoded by a linker or PCR primer) may be inserted between the selected polypeptide and the modification recognition (HMK) site. In many cases, however, only 17 amino acids will be fused to the selected polypeptide.

-26-

pFEK-plasmids can direct the expression of FLAG-fusion proteins (i.e., fusion proteins having a FLAG epitope affinity ligand) from the T7 polymerase promoter in an appropriate host (e.g., a bacterial cell capable of constitutive or inducible expression of the T7 polymerase; for examples of suitable hosts and induction protocols, see Studier et al., Meth. Enzymol. 185: 60-89 (1990)). The construction of several pFEK-plasmids is described in Example 5. The encoded fusion proteins were expressed in a bacterial host, and a lysate was prepared. Partially purified fractions containing each fusion protein were subjected to phosphorylation with the catalytic subunit of the cAMP-dependent protein kinase using [ $\gamma$ - $^{32}$ P]ATP as a (labeled) modification donor. The FEK-fusion proteins were labeled to high specific activity to give [ $^{32}$ P]-labeled FEK-fusion proteins (which are also [ $^{32}$ P]-labeled FLAG-fusion proteins).

Note that vectors of the present invention related to pGTK- and pFEK- vectors can be constructed lacking the cleavable linker sequence. These vectors would have a pGK- or pFK- prefix. A fusion protein comprising a selected polypeptide expressed from a pGK- or pFK-vector is referred to as pGK-fusion protein or pFK-fusion protein, respectively.

A variety of commercially available vectors comprising an affinity ligand and cleavable linker are available which can be modified by the introduction of a modification site (e.g., a

-27-

phosphorylation site), and where required, one or more restriction sites for the in-frame insertion of a selected polypeptide using recombinant DNA techniques. For example, (1) the pGEX-3X vector  
05 (Pharmacia) comprising a GST affinity ligand and Factor Xa cleavable linker, (2) the pMAL-C vector (Biolabs) comprising a malB affinity ligand and Factor Xa cleavable linker, (3, 4) the protein A gene fusion expression vectors, pRIT2T and pRIT5  
10 (Pharmacia), comprising a protein A affinity ligand, (5, 6) as well as pDS and pQE-vectors (Qiagen), comprising a histidine hexamer affinity ligand, could be modified in this way. The vector pRIT2T, vector pRIT5, pDS and pQE-vectors, could be further modified  
15 by the insertion of a sequence encoding a cleavable linker. Convenient affinity matrices for fusion proteins encoded by the foregoing were discussed above.

#### Methods of Producing a Modified Fusion Protein

20 The present invention further relates to methods for producing a modified fusion protein. In particular, a rapid and convenient method for labeling a fusion protein using a radiolabel or non-radioactive label donated by [ $\gamma$ -labeled]ATP is  
25 provided. In addition, the methods used are mild; a feature which is important for preservation of the structure and biological activity (e.g., binding, antigenicity, activity) of the components of the fusion protein, and of the selected polypeptide in  
30 particular.

-28-

For example, fusion proteins of the present invention comprising an affinity ligand portion and a selected polypeptide portion are expressed in a host cell carrying a vector of the present invention. The  
05 host cells are propagated under conditions which permit expression of the vector. For example, expression from the particular pGTK- and pFEK- vectors described in the Examples requires induction by IPTG. The host cells are lysed using known  
10 techniques to obtain a lysate containing the fusion protein.

In one embodiment, the lysate can be crudely fractionated as in Example 5 and directly labeled in the presence of [ $\gamma$ -labeled]ATP (e.g., [ $\gamma$ -<sup>32</sup>P]ATP),  
15 and a cAMP-dependent protein kinase, such as the catalytic subunit of the cAMP-dependent protein kinase from bovine heart muscle. A suitable formulation for a kinase reaction buffer and reaction stop buffer is given below.

20 In another embodiment, the fusion protein is modified (e.g., phosphorylated) while bound to an affinity matrix. The fusion protein present in the lysate is captured on an affinity matrix. This step is carried out by contacting the lysate with an  
25 appropriate affinity matrix (i.e., an affinity matrix comprising a specific binding partner of the affinity ligand present in the fusion protein), under conditions which permit binding of the affinity ligand portion of the fusion protein to the affinity  
30 matrix. Suitable conditions for a variety of affinity ligands and matrices are known in the art.

-29-

For example, GST-fusion proteins (e.g., a GTK-fusion) can be captured on immobilized glutathione as an affinity matrix. As shown in Example 2, the GST-fusion protein can be captured on the affinity matrix in the presence of a wash buffer, (which was also used as the lysis buffer). The components of the wash buffer permit binding of the GST portion of the fusion proteins to the matrix via the specific binding partner (glutathione). In one embodiment, the wash buffer comprises a buffer such as Tris, salt (e.g., NaCl), a chelator (e.g., EDTA), and a non-ionic detergent such as nonidet P-40.

As discussed above, FEK-fusion proteins can be captured on an anti-FLAG antibody affinity matrix. When the FLAG peptide is internal (e.g., not immediately at the N-terminus, as in pAR( $\Delta$ RI)59/60), anti-FLAG M2 antibody can be used as the specific binding partner. Suitable conditions and formulations (e.g., for wash buffer and elution buffer) for anti-FLAG M2 affinity chromatography can be obtained from International Biotechnologies, Inc.

Once bound, the fusion proteins can then be washed in order to remove unbound material (e.g., contaminating proteins other than the fusion protein). The wash buffers described above are suitable for this purpose.

The affinity matrix with bound fusion protein attached is then equilibrated (contacted) with a reaction buffer suitable for the modification reaction. In a phosphorylation reaction, for example, a kinase buffer such as HMK buffer is

-30-

selected. The particular kinase reaction buffer will be adapted to the kinase selected. Suitable reaction buffers for a variety of characterized kinases are available. HMK reaction buffer, suitable for use  
05 with the cAMP-dependent protein kinase from bovine heart muscle (HMK), comprises (1) a buffering agent such as Tris, (2) a reducing agent such as dithiothreitol (DTT), (3) a salt such as NaCl, and (4) a source of magnesium ions or other appropriate  
10 ion, such as  $MgCl_2$ . Although the DTT can be added separately, it is present in the kinase reaction buffer. In addition, a modification enzyme and modification donor are added. In one particular phosphorylation reaction, a protein kinase preparation  
15 comprising the HMK enzyme is added, and [ $\gamma$ -labeled]ATP. Typically, the protein kinase will be diluted in a suitable solution prior to addition. In the examples, a variety of fusion proteins were efficiently radiolabeled by phosphorylation using HMK  
20 reaction buffer, the catalytic subunit of the HMK and [ $\gamma$ - $^{32}P$ ]ATP as a modification donor.

The conditions appropriate for phosphorylation of the bound fusion protein to produce a bound labeled (e.g., radiolabeled) fusion protein will vary  
25 with the protein kinase selected. Typically, the steps subsequent to culturing the host cell are carried out at 4 °C. However, phosphorylation reactions with the catalytic subunit of the cAMP-dependent protein kinase from bovine heart  
30 muscle, have previously been carried out at 37 °C. In the present method, phosphorylation can be carried

-31-

out at 37 °C. However, in another embodiment of the present invention, efficient phosphorylation is carried out at 4 °C. This temperature can preserve the activity of fusion proteins, and of the selected polypeptide portion in particular.

Optionally, the reaction can be quenched by addition of a stop buffer comprising a component capable of inhibiting the modification reaction. The component capable of inhibiting the reaction will vary with the modification enzyme, but can include inhibitors (e.g., reaction products, competitors), chelators, or other agents which stop the reaction. It can be desirable to stop a reaction when, for example, in subsequent steps, undesired modifications can occur. However, wash steps can serve to remove modification enzymes.

For example, in a phosphorylation reaction with cAMP-dependent protein kinase, an HMK stop buffer can be added, comprising a buffer such as sodium phosphate, a reaction inhibitor such as sodium pyrophosphate, and a chelator such as EDTA. HMK stop buffer can optionally contain a carrier such as bovine serum albumin or glycerol. As a reaction product of a phosphorylation reaction, sodium pyrophosphate inhibits the kinase reaction. In addition, chelating  $Mg^{2+}$  ions inhibits the reaction.

An advantage of the present method is that unincorporated label is easily removed by washing the affinity matrix with bound labeled fusion protein. The wash buffers described above, which permit binding of the affinity ligand to the affinity matrix

-32-

via the specific binding partner can be used in this step. Other methods for labeling proteins with the HMK enzyme have relied upon extensive dialysis to remove unincorporated label (e.g., Zhao, X.-X. et al., Analyt. Biochem. 178: 342-347 (1989)). The removal of unincorporated label by washing is advantageous because extensive dialysis can compromise the function of some proteins and results in the production of large quantities of contaminated dialysis buffer.

If desired, the modified fusion protein may be isolated for use by releasing the fusion protein from the affinity matrix by washing with a suitable elution buffer. Elution buffers were discussed above. For example, a labeled GST-fusion protein such as a GTK-fusion protein can be eluted from immobilized glutathione with an elution buffer comprising glutathione (e.g., reduced glutathione). In one embodiment, an elution buffer comprising reduced glutathione, a buffer such as Tris, and a salt such as NaCl is used to release labeled GST-fusion proteins from the affinity matrix.

In the case where a cleavable linker is present, after elution, the fusion protein can be cleaved in vitro to free the modified (e.g., labeled) polypeptide portion. Cleavage is accomplished by contacting the fusion protein with an appropriate specific protease under conditions which permit the cleavage reaction. Such conditions for cleavage by enterokinase, thrombin and factor Xa, for example, are known. The affinity ligand portion or any

-33-

uncleaved product can be removed by adsorption on the appropriate affinity matrix in the presence of wash buffer for example. The modified or labeled selected polypeptide portion can be recovered in the  
05 supernatant.

Optionally, prior to elution from the affinity matrix, the modified polypeptide portion can be released from the affinity ligand portion by cleaving the modified fusion protein at the cleavable linker.  
10 An example protocol for cleavage by a specific protease on a column is provided in Abath, F. and A. Simpson, Biotechniques 10: 178 (1991), the teachings of which are herein incorporated by reference.

Smith (EP 0,293,249; PCT/AU88/00164; NZ 224,663)  
15 and Smith and Johnson (Gene 67: 31-40 1988)) also describe conditions for capture of GST-fusion proteins on immobilized glutathione, washes, elution and cleavage protocols.

In one embodiment of the present invention, a  
20 kit for preparing a labeled fusion protein is provided, comprising (1) an affinity matrix to which a portion of the fusion protein can bind, (2) a wash buffer which permits binding of the fusion protein to the affinity matrix, (3) a modification reaction  
25 buffer such as a protein kinase reaction buffer, and (4) a modification enzyme preparation such as a protein kinase preparation. Optionally, the kit can contain a reaction stop buffer or an elution buffer comprising a release component. The nature of these  
30 elements has been explained above in more detail.

In another embodiment, a kit can comprise a vector of the present invention (e.g., pGEX-2TK or

-34-

PAR[ $\Delta$ RI]59/60). The vector can be present in a suitable host cell or free from the host cell. Furthermore, the kit comprises each of the elements (1) through (4) of the kit described above, and if  
05 desired, a reaction stop buffer. In other embodiments, a modification donor such as  $\gamma$ -labeled ATP can also be included in the kit.

In a phosphorylation reaction, a protein kinase preparation is provided comprising a protein kinase.  
10 In the protein kinase preparation, the kinase may be in the form of a precipitate or solution. The kinase can be dissolved or diluted as necessary prior to use.

The affinity matrix can be present in a suitable  
15 containing means (e.g., a vial or cartridge). In one embodiment, the affinity matrix is provided in a cartridge unit, the combination of which acts as a column. Various components (e.g., the lysate, wash buffers, label) can be inserted into the cartridge  
20 and passed through the affinity column to facilitate the labeling procedure. In one embodiment, immobilized glutathione can be incorporated into a cartridge unit for capture of GST-fusion proteins.

#### 25 Uses of the Present Invention

Vectors of the present invention can direct the expression of large amounts of a fusion protein in a host cell. These fusion proteins can be conveniently  
30 labeled using the methods described, and have many uses in research, diagnostic and therapeutic applications.

-35-

For example, radiolabeled proteins are useful in imaging. In one embodiment, radiolabeled fusion protein or radiolabeled selected polypeptide portion can be used to locate cells carrying a ligand  
05 recognized by the selected polypeptide (e.g., a receptor, cell surface component, or antigen). For example, a growth factor can be used as the selected polypeptide, and a labeled fusion protein of the present invention comprising the growth factor can be  
10 used to detect cells carrying a receptor for the growth factor.

Similarly, incorporation of a suitable radiolabel permits the use of fusion proteins of the present invention in targeted radiotherapy. In other  
15 words, the selected polypeptide (e.g., a hormone, growth factor, antibody) present in the fusion protein can seek out specific cells and damage or destroy them due to the attached radiolabel. An isotope with the desired energy and half-life can be  
20 selected for this purpose. Other labels incorporated by methods of the present invention and capable of killing cells can also be used.

For example, a labeled antibody or portion thereof (e.g.,  $F_v$ , Fab, etc.) produced by the method  
25 of the present invention could be used in this manner. Such fusion proteins would be useful for many other applications as well. The term antibody as used herein refers to antibodies such as chimeric antibodies, single chain antibodies, bifunctional  
30 antibodies and other antibody variants, including individual chains (e.g., a heavy chain). A variety

-36-

of methods for expressing immunoglobulins are available (see e.g., Skerra, A. and A. Pluckthun, Science 240: 1038-1041 (1988)).

In one embodiment, antibody chains are expressed  
05 as fusion proteins of the present invention in E. coli. A signal sequence (e.g., ompA, phoA) is fused to the N-terminus of an antibody chain, which is followed by at least one modification recognition sequence, an optional cleavable linker, and an  
10 affinity ligand. On expression in E. coli, the signal sequence is cleaved from the fusion protein, freeing the amino terminus of the antibody chain, and preserving the binding function of the variable region. Fusion genes encoding both chains can be  
15 expressed in this manner from two vectors or a single vector encoding two fusion genes. Alternatively, either the heavy or light chain could be expressed as a fusion protein of the present invention and the complementary chain could be expressed in the same  
20 cell using typical antibody expression vectors.

In addition, as shown herein, labeled fusion proteins of the present invention can be used as probes. As shown in Example 4, labeled fusion proteins such as <sup>32</sup>P-GTK-fusion proteins, can be used  
25 as probes for screening expression libraries for proteins capable of binding the selected polypeptide. For example, a fusion protein of the present invention comprising a selected polypeptide is expressed, purified by virtue of the affinity ligand,  
30 and specifically labeled at the modification recognition sequence. The labeled, homogeneous

-37-

fusion protein is then incubated with filters on which proteins expressed by individual plaques of a target library have been immobilized. Following this hybridization step, the filters are washed, and  
05 processed for identification (i.e., by detection of the label) of plaques which produce recombinant proteins capable of specifically interacting with the labeled fusion protein probe.

In particular, a cDNA encoding a portion (the E7  
10 binding domain) of the retinoblastoma susceptibility gene product (pRB) was prepared as a GTK-fusion protein and used to screen a library for proteins which interact with pRB. A number of positive clones were isolated. One of these clones, designated  
15 RBAP1, displays properties expected of a cellular protein capable of interacting with pRB. Labeled fusion proteins of the present invention comprising a selected polypeptide can also be used as probes in Western blot formats (Example 4).

20 In addition, as shown in Example 4, labeled fusion proteins of the present invention can be used to identify proteins present in complex mixtures (e.g., whole cell extracts), which are capable of interacting with the selected polypeptide. The  
25 fusion proteins can be used in protocols to isolate the interacting proteins. For example, the interacting proteins can be captured by the selected polypeptide portion of a labeled fusion protein which is bound to an affinity matrix. Subsequent cleavage  
30 at a cleavable linker could release the complex of the interacting protein and a labeled selected

-38-

polypeptide portion for further analysis.

Alternatively, a purified labeled fusion protein probe can also be used to determine the amount of or to detect an interacting protein or other substance  
05 in a complex mixture using immunoassay techniques, for example.

In one embodiment, the binding of a specifically interacting protein (e.g., an antigen, hormone) or other substance to an immobilized (e.g., bound to an  
10 affinity matrix) fusion protein of the present invention can induce a change in the attributes of the bound fusion protein. Other substances include, but are not limited to agents such as a drug, toxin, substrate or other ligands capable of interacting  
15 with the selected polypeptide portion of the fusion protein. The difference between the unbound versus the bound state could be taken advantage of, as for example, in an assay format. For example, the presence of the interacting protein or other  
20 substance in a complex mixture (e.g., blood) could be monitored. For example, binding of an interacting protein could induce a conformational change that shields the labeling (e.g. phosphorylation) site. The complex mixture containing an interacting protein  
25 is contacted with the fusion protein, which is bound to the affinity matrix. The combination is washed, equilibrated with a suitable reaction buffer and subjected to the labeling procedure. The extent of labeling is compared to a control which lacks the  
30 interacting polypeptide and the presence of the interacting protein in the complex mixture is

-39-

indicated by inhibition of labeling. By comparison with the degree of inhibition achieved by a standard, such a procedure can also provide information regarding the quantity of interacting protein which  
05 is present.

The vectors and methods of the present invention can facilitate the characterization of cDNA clones. For example, an uncharacterized cDNA can be inserted into a vector of the present invention allowing rapid  
10 purification of the protein, without specific knowledge of its properties. The protein can be labeled with the maintenance of biological activity, and the labeled fusion protein can be used to detect potential interactions with other cellular proteins.

15 The present invention will now be illustrated by the following Examples, which are not intended to be limiting in any way.

#### Introduction to Examples 1-4

Recently it has been demonstrated that the  
20 adenovirus E1A protein, SV40 and polyomaviral T antigens, and the human papilloma virus E7 protein can bind to the retinoblastoma susceptibility gene product, pRB. Neither these viruses nor these proteins are thought to be closely related to one  
25 another, yet each of these proteins shares a short, homologous, colinear, transforming sequence, typified by E1A conserved region 2 (CR2), which appears to be responsible for high affinity binding to pRB. The region of pRB which interacts with this motif has

-40-

been mapped. To date, all spontaneously occurring, loss of function, pRB mutations which do not grossly compromise pRB stability, map to this region. This led to the hypothesis that pRB, in the course of  
05 regulating cell growth, must interact with one or more cellular proteins bearing a sequence structurally resembling the viral pRB-binding motif. If such a model were correct, one would predict that pRB inactivation might be achieved either as a result  
10 of competing complex formation with one of the above viral oncoproteins, or as a result of RB mutations affecting the T/E1A/E7-binding domain.

-41-

Example 1Construction of pGEX-2TK

Two synthetic oligonucleotides were designed, which, on annealing, encode a cAMP-dependent protein kinase recognition site and three restriction sites. Codon utilization for the synthetic duplex was based on the prokaryotic codon utilization data of (Grantham, R. et al., Nucleic Acids. Res. 9: r43-r74 (1981)).

The oligonucleotides were synthesized using standard techniques and were annealed in vitro (Gait, M.C.J., (1984) Oligonucleotide Synthesis: A Practical Approach, (I.R.L. Press; Oxford)). The resulting duplex has 5'-overhangs compatible with BamHI- and EcoRI-cut DNA and is shown below:

ArgArgAlaSerVal  
5'-GATCTCGTCGTGCATCTGTTGGATCCCCGGG  
AGCAGCACGTAGACAACCTAGGGGCCCTTAA-5'

Plasmid pGEX-2T (Pharmacia) was linearized with BamHI and EcoRI; and the vector fragment was ligated to the synthetic duplex to make plasmid pGEX-2TK. Incorporation of the duplex into pGEX-2T was confirmed by restriction and DNA sequence analysis. DNA sequencing was performed using a Sequenase 2.0 kit (United States Biochemical Corp.) with the protocol provided by the manufacturer. The structure of pGEX-2TK in the region surrounding the kinase site is shown in Figure 1.

-42-

As can be seen in Figure 1, the incorporation of the synthetic duplex DNA into pGEX-2T resulted in the insertion of codons for a cAMP-dependent protein kinase recognition sequence, Arg-Arg-Ala-Ser-Val, immediately downstream of the thrombin recognition site present in parent plasmid pGEX-2T. In addition, the synthetic duplex encoded a multiple cloning site (MCS) downstream of the kinase recognition sequence, such that the insertion of the duplex led to the regeneration of the MCS of the parent plasmid, with restriction sites for BamHI, SmaI, and EcoRI. As in the parent plasmid, the MCS is followed by a sequence with stop codons in all three reading frames.

### Example 2

#### 15      Generation of <sup>32</sup>P-GST Fusion Proteins

##### Construction of pGEX-2TK Vectors Encoding RB Proteins

RB cDNAs encoding residues 379-792 or residues 379-928 of the retinoblastoma susceptibility gene product, both of which span the T/E1A-binding region, were subcloned into pGEX-2TK. In addition, an RB cDNA encoding residues 379-928, having a naturally occurring, loss of function, RB point mutation (379-928;706F), which is known to abrogate T/E1A binding, was subcloned into pGEX-2TK. These pGEX-2TK recombinants are named pGTK-RB(379-792), pGTK-RB(379-928) and pGTK-RB(379-928;706F), respectively. The encoded fusion proteins are named GTK-RB(379-792), GTK-RB(379-928) and

-43-

GTK-RB(379-928;706F), respectively. These RB fusion proteins are also referred to as GST-RB fusion proteins or RB fusion proteins.

Two of the RB cDNAs had been generated in an earlier study by Kaelin et al. (Kaelin, W.G. et al., Mol. Cell. Biol. 10(7): 3761-3769 (1990)), incorporated herein by reference, using the polymerase chain reaction. The RB cDNA inserted into pGEX-2TK to make pGTK-RB(379-928) corresponds to RB deletion RB d1 1-378 (Kaelin, W.G. et al., Mol. Cell. Biol. 10(7): 3761-3769 (1990)). The RB cDNA inserted into pGEX-2TK to make pGTK-RB(379-792) corresponds to RB deletion RB d1 1-378;793-928. Residue 379 is a methionine residue.

Each ampimer used to generate the cDNAs contained a BamHI site, such that the resulting PCR product, upon digestion with BamHI, could be ligated in frame into the unique BamHI site in pGEX-2T (Pharmacia). The 3' ampimer contained a TGA stop codon as well. The PCR fragments encoding these two cDNAs were each cleaved with BamHI and cloned into pGEX-2T, which had been linearized with BamHI, and treated with calf intestinal phosphatase. The resulting constructs were named pGT-RB(379-928) and pGT-RB(379-792). The cDNAs were cleaved from the latter constructs using BamHI, and were each subcloned into pGEX-2TK, which had been cleaved with BamHI. The resulting plasmids are named pGTK-RB(379-928) and pGTK-RB(379-792). Two additional amino acids were incorporated into the sequence at the BamHI site due to the structure of

-44-

the PCR primers used to clone both cDNAs, such that the amino acid sequence at the junction of the phosphorylation (kinase) site and the first RB residue (Met<sub>379</sub>) is:

05 (-Arg-Arg-Ala-Ser-Val-Gly-Ser-Ala-Thr-Met<sub>379</sub>-).

The mutant RB cDNA and referred to herein as either RB(379-928;706F) or RB(379-928;pm706) corresponds to the cDNA designated (379-928;pm 706) constructed by Kaelin et al. (Kaelin, W.G. et al.,  
10 Cell 64: 521-532 (1991)), also incorporated herein by reference. The mutant cDNA fragment was generated by PCR of reverse-transcribed mRNA from cells containing the mutation. The PCR product was cleaved at the  
15 unique NcoI and BsmI sites in the RB gene (which span the 706F mutation), and ligated into pGT-RB(379-928), from which the wild type RB cDNA NcoI-BsmI segment had been excised. The resulting plasmid, pGT-RB(379-928;706F), encoded the 379-928;706F cDNA. The plasmid was cleaved with BamHI to release the  
20 cDNA, and the cDNA fragment was subcloned into pGEX-2TK which had been cleaved with BamHI to make pGTK-RB(379-928;706F). The amino acid sequence at the junction of the phosphorylation (kinase) site and the first RB residue (Met<sub>379</sub>) is also  
25 (-Arg-Arg-Ala-Ser-Val-Gly-Ser-Ala-Thr-Met<sub>379</sub>-).

The products encoded by all three cDNAs have been tested previously for their ability to bind to SV40 T antigen, the adenovirus E1A gene product, and putative cellular RB-binding proteins, when expressed  
30 as pGEX-2T encoded GST fusion proteins. The fusion proteins encoded by pGT-RB(379-928) and

-45-

pGT-RB(379-792) were able to bind to T antigen, E1A, and the putative cellular RB-binding proteins, while the mutant protein encoded by pGT-RB(379-928;706F) did not (Kaelin, W.G. et al., Cell 64: 521-532 (1991)).

#### Expression and Purification of GST-Fusion Proteins

The expression of the pGEX-2TK-encoded protein and recombinant pGEX-2TK-encoded GST fusion proteins in E. coli (DH5 $\alpha$ ; Bethesda Research Laboratories), and the subsequent recovery of the proteins on glutathione sepharose was carried out essentially as described by Smith and Johnson (Smith, D.B. and Johnson, K.S., Gene 67: 31-40 (1988); Kaelin, W.G. et al., Cell 64: 521-532 (1991)). Fresh overnight cultures of E. coli DH5 $\alpha$ , transformed with either pGEX-2TK or pGEX-2TK recombinants, were diluted 1:10 in Luria-Bertani (LB) medium containing ampicillin (100  $\mu$ g/ml) and incubated for a total of 5 hours at 37  $^{\circ}$ C, with shaking. After 1.0 hour of growth at 37  $^{\circ}$ C, isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG; Bethesda Research Laboratories) was added to a final concentration of 0.1 mM.

For analysis of total bacterial protein content, aliquots of each bacterial culture were pelleted in a microcentrifuge, were boiled in urea-SDS cracking buffer (0.01 M sodium phosphate [pH 7.2], 1%  $\beta$ -mercaptoethanol, 1% SDS, and 8M urea), and were loaded onto an SDS-polyacrylamide gel. Proteins were visualized by Coomassie blue staining.

For phosphorylation of protein and/or protein

-46-

recovery using glutathione sepharose (Pharmacia), bacterial cultures were pelleted at 5000 x g for 5 min at 4 °C, and resuspended 1/10 the original culture volume of NETN (20 mM Tris [pH 8.0], 100 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40). The bacteria were then lysed on ice by mild sonication and subjected to centrifugation at 10,000 x g for 5 min at 4 °C. The clarified bacterial sonicates, containing the pGEX-2TK-encoded protein or relevant GST fusion protein, were rocked for 15-30 minutes at 4 °C with glutathione sepharose (20-30  $\mu$ l/ml bacterial sonicate). The glutathione sepharose beads (Glutathione Sepharose 4B, Pharmacia) had been washed three times and resuspended 1:1 v/v in NETN (see above) supplemented with 0.5% non-fat powdered milk prior to use. This step and subsequent steps were also carried out at 4°C.

#### Phosphorylation of Proteins

The sepharose beads, with bound protein, were then washed three times with NETN followed by one wash with 1X HMK buffer (20 mM Tris pH [7.5], 100 mM NaCl, 12 mM  $MgCl_2$ ). The supernatant was aspirated with a 23G needle and the sepharose was resuspended in 2-3 bead volumes of 1X HMK buffer containing 1 unit/ $\mu$ l of the catalytic subunit of cAMP-dependent protein kinase (Sigma Chemical Co., St. Louis, MO), 1  $\mu$ Ci/ $\mu$ l  $^{32}$ P- $\gamma$ ATP (6000 Ci/mMole, 10 mCi/ml, New England Nuclear), and 1 mM DTT. The kinase reaction was allowed to proceed for 30 minutes at 4 °C, with periodic agitation of the sepharose to maintain a

-47-

suspension. The reaction was terminated by the addition of 1.0 ml of HMK stop buffer (10 mM Na Phosphate [pH 8.0], 10 mM Na Pyrophosphate, 10 mM EDTA, 1 mg/ml bovine serum albumin). Following a  
05 brief spin in a microcentrifuge, the supernatant was again removed using a 23 G needle and the sepharose was washed 5 times with NETN to remove any unincorporated label. Incorporation of label can be determined while the protein is attached to the bead.

#### 10 Elution With Reduced Glutathione

For further studies, the fusion protein can be eluted from the beads. After the final wash the residual supernatant was aspirated using a 23 G needle and the labeled or unlabeled GST fusion  
15 protein was eluted by rocking the sepharose for 10-15 minutes in 10-50 bead volumes of 20 mM reduced glutathione, 100 mM Tris [pH 8.0], 120 mM NaCl.

#### Effect of HMK Sequence on $^{32}\text{P}$ -Incorporation In Vitro

The ability of the Arg-Arg-Ala-Ser-Val sequence  
20 to convert the GST fusion proteins into substrates for the catalytic subunit of cAMP dependent protein kinase was tested. GST fusion proteins encoded by the pGEX-2TK recombinants, as well as the corresponding pGEX-2T constructs, were overexpressed  
25 in bacteria and recovered on glutathione sepharose as described above. As described in more detail above, the purified fusion proteins, while still non-covalently bound to glutathione sepharose, were incubated with  $^{32}\text{P}$ - $\gamma$ -ATP and the purified catalytic

-48-

subunit of cAMP-dependent protein kinase for 30 min at 4°C, after which a kinase stop buffer was added. The sepharose, with bound protein, was extensively washed to remove unincorporated label, transferred to  
05 clean eppendorf tubes, and subjected to Cerenkov counting.

The GST-RB fusion proteins were readily phosphorylated in vitro, provided that the kinase recognition sequence was present. Figure 2 shows the  
10 effect of inserting an HMK sequence on  $^{32}\text{P}$  incorporation (cpm) in vitro data on four GST-RB fusion proteins. In particular, GST-RB(379-792) and GST-RB(379-792;pm706) fusion proteins which were encoded by pGEX-2TK and which contained an HMK  
15 sequence (hatched bars), showed significant incorporation of  $^{32}\text{P}$ . In contrast, GST-RB(379-792) and GST-RB(379-792;pm706) fusion proteins which were encoded by pGEX-2T and lacked an HMK sequence (solid bars), were not significantly radiolabeled.

20 In another experiment, fusion proteins from 10 ml of bacterial culture were recovered on 25  $\mu\text{l}$  of glutathione sepharose beads, and were subjected to phosphorylation using the same protocol. A GST-RB(379-928) fusion protein expressed from  
25 pGEX-2TK and having a kinase site was labeled with  $^{32}\text{P}$  ( $3 \times 10^6$  cpm), while the GST-RB(379-928) fusion protein expressed from pGEX-2T and lacking a kinase site was not comparably phosphorylated ( $1 \times 10^4$  cpm). Similarly, a GST-RB(379-928);pm706 fusion protein  
30 expressed from pGEX-2TK and having a kinase site was labeled with  $^{32}\text{P}$  ( $3 \times 10^6$  cpm), while the GST-RB(379-928);pm706 fusion protein expressed from pGEX-2T and lacking a kinase site was not comparably phosphorylated ( $1 \times 10^4$  cpm).

-49-

Example 3Structural Integrity of Fusion ProteinsExpression of Fusion Proteins in E. coli

As discussed above, the GST-RB fusion proteins  
05 expressed from pGEX-2T-derived plasmids  
pGT-RB(379-928) and pGT-RB(379-792) were able to bind  
to T antigen, ElA, and the putative cellular  
RB-binding proteins, while the mutant fusion protein  
encoded by pGT-RB(379-928;706F) did not (Kaelin, W.G.  
10 et al., Cell 64: 521-532 (1991)). This observation  
suggests that the binding function and structure of  
the RB proteins is not grossly disturbed when they  
are expressed as part of a fusion protein with GST.

It was further determined that the insertion of  
15 the Arg-Arg-Ala-Ser-Val (RRASV) sequence did not  
grossly alter bacterial expression of GST-RB fusion  
proteins expressed from pGEX-2TK. For this  
determination, whole cell lysates of IPTG-induced  
cultures were prepared and total bacterial protein  
20 content was analyzed directly on SDS-polyacrylamide  
gels as described above. Protein from E. coli  
transformed with pGTK-RB(379-928), pGTK-RB(379-792),  
or pGTK-RB(379-928;pm706) were analyzed. The band  
intensity for each fusion protein was comparable to  
25 that for the corresponding construct lacking the  
phosphorylation site. Each fusion protein had an  
apparent molecular weight consistent with the size of  
the pRB fragment encoded by the RB cDNA insert and  
the 26 kD GST polypeptide.

-50-

### Purification of Fusion Proteins

The subsequent purification of GST-RB fusion proteins expressed from pGEX-2TK was not altered by the insertion of the RRASV sequence. Bacterial  
05 sonicates of each of the transformants were also prepared, and proteins were subjected to glutathione sepharose chromatography as described above. The sepharose beads, with bound protein, were then washed three times with NETN, and the bound proteins were  
10 eluted by boiling in SDS sample buffer (2% SDS, 10 % glycerol, 62 mM Tris [pH 6.8]). The supernatant was subjected to electrophoresis on a 10% polyacrylamide gel and visualized by Coomassie blue staining. After purification, the intensity of the bands for  
15 GTK-fusion proteins expressed in E. coli transformed with pGTK-RB(379-928), pGTK-RB(379-792), or pGTK-RB(379-928;pm706) was comparable to that observed with for the corresponding constructs lacking the kinase site (i.e., pGT-RB(379-928),  
20 pGT-RB(379-792), or pGT-RB(379-928;pm706), respectively)

The following materials and methods were used in subsequent experiments.

### Cells and Culture Conditions

25 The retinoblastoma cell line WERI-Rb27 (a gift of Dr. Wen-Hwa Lee) Huang, H.-J. et al., Science 242: 1563-1566 (1988)) and 293 cells, a human embryonic cell line transformed by a fragment of the Adenovirus 5 genome (Graham, F.L. et al., J. Gen. Virol. 36:

-51-

59-72 (1977)), were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal calf serum (Gibco). Akata cells were grown in RPMI or DMEM, with 10% fetal calf serum (Gibco). C57Bl/6 primary  
05 mouse embryo fibroblasts (MEF) lines expressing either wild-type (MEF<sup>+</sup>Tex) or mutant (MEF<sup>+</sup>K1) forms of T antigen were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal calf serum (Gibco) and G418 (150 ug/ml) (Ewen, M.E. et al., Cell  
10 58: 257-267 (1989)). All cells were grown at 37°C in a humidified, 10% CO<sub>2</sub>-containing atmosphere. Radioisotopic labelling of cells and preparation of cell lysates was as described previously (Kaelin, W.G. et al., Cell 64: 521-532 (1991)).

#### 15 Antibodies

Tissue culture supernatants were the source of monoclonal antibodies PAb 419 and M73 (Harlow, E. et al., J. Virol. 39: 861-869 (1981); Harlow E. et al., J. Virol. 55: 533-546 (1985)). The use of these  
20 antibodies for immunoprecipitation and western blotting was described previously (Kaelin, W.G. et al., Cell 64: 521-532 (1991)), except that electrophoretic transfer of proteins to nitrocellulose was performed without the addition of methanol to the  
25 transfer buffer.

#### Binding of Fusion Proteins to T Antigen and ElA

It was also determined that the insertion of the RRASV sequence did not demonstrably alter the binding behavior of GST-RB fusion proteins, GTK-RB(379-928),

-52-

GTK-RB(379-792), and GTK-RB(379-928;706F), expressed from pGEX-2TK. The pRB binding assays for T antigen and E1A binding were performed essentially as described by Kaelin et al. (Kaelin et al., Cell 64: 521-532 (1991)).

Fusion proteins GTK-RB(379-928) and GTK-RB(379-792) (bound to glutathione-sepharose) retained the ability to bind to SV40 T antigen or adenovirus E1A in solution. In contrast, no significant binding of T antigen or E1A to the mutant RB fusion, GTK-RB(379-928;706F), or GST expressed from pGEX-2TK was observed under the same conditions.

#### Integrity of Phosphorylated Fusion Proteins

To determine whether the kinase reaction led to a significant alteration in the RB T/E1A-binding region of the GTK-RB fusion proteins, the following experiment was performed. Whole cell lysates containing wild-type T antigen (MEF $\Psi$ Tex cells), the RB-binding and transformation defective T antigen mutant K1 (MEF $\Psi$ K1 cells), or E1A (293 cells) were resolved by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose filters (See protocols in Example 4). GST-RB fusion proteins were kinased in vitro and eluted from the glutathione sepharose in the presence of reduced glutathione as described above. The eluted labeled protein was incubated overnight with nitrocellulose strips cut from the filters. The filters were then washed and subjected to autoradiography. Hybridization conditions and washes were as described below in Example 4, Hybridization of Filters.

-53-

Similar to the unphosphorylated version,  
<sup>32</sup>P-GTK-RB(379-792) bound to E1A and wild-type T from  
MEFvTex cells, but not to the T mutant K1 from MEFvK1  
cells. The presence of equivalent amounts of T and  
05 K1 in this assay was confirmed by immunoblotting with  
a monoclonal antibody directed against T antigen.

The binding of <sup>32</sup>P-GTK-RB(379-792) protein to  
E1A was inhibited by the presence of a synthetic  
peptide replica of the human papillomavirus E7  
10 RB-binding motif (wild-type E7 residues 16-32). In  
contrast, a point mutant derivative of this peptide  
with a glu to gln change at residue 26, which is  
known to abrogate in vitro RB binding, was inert as  
an inhibitor. In addition, <sup>32</sup>P-GTK-RB(379-928), but  
15 not <sup>32</sup>P-GTK-RB(379-928);706F), exhibited E1A binding  
in this assay. Thus, it appeared that the structural  
integrity of the T/E1A-binding regions in the GST-RB  
chimeras was preserved during the kinase procedure  
and subsequent elution.

20 The experiments described in this example  
indicate that the expression, behavior during  
purification and structural integrity of different  
polypeptides inserted into pGEX-2TK is not grossly  
affected by insertion of a kinase sequence or by  
25 subsequent phosphorylation.

-54-

Example 4Screening for Proteins Capable of Specific  
Interaction with the Protein of Interest and  
Isolation of Genes Encoding Such Proteins05 Library Manipulations

For primary screening, libraries were plated at approximately 40,000 pfu/150 mm plate x 30 plates. Induction of  $\beta$ -galactosidase fusion protein expression with IPTG impregnated nitrocellulose filters was performed as described by Singh *et al.* (Biotechniques 7: 252-261 (1989)). Additional library manipulations including screening with  $^{32}\text{P}$ -labelled cDNA probes, purification of specific clones, preparation of recombinant phage DNA, and subcloning of cDNAs into the sequencing vector pBKS (Stratagene) were performed using standard techniques.

Preparation of Nitrocellulose Filters for  
Hybridization

20 Electrophoretic transfer of proteins from SDS-polyacrylamide gels to nitrocellulose for western blot analysis was carried out in 192 mM glycine, 25 mM Tris(Base), and 0.01% SDS. Plaque lifts were performed as described above. The nitrocellulose filters were placed, directly into 1X HBB (25 mM  
25 Hepes-KOH[pH7.7], 25 mM NaCl, 5mM  $\text{MgCl}_2$ , 1 mM DTT) supplemented with 5% non-fat powdered milk and 0.05% NP-40 without drying and incubated overnight. This

-55-

and subsequent manipulations were performed at 4°C with gentle rocking. The filters were then denatured and renatured as described by Vinson et al. (Genes and Dev. 2: 801-806 (1988)). Processing of multiple  
05 filters was done in batch.

Following renaturation, the filters were placed in fresh 1X HBB supplemented with 5% non-fat powdered milk and 0.05% NP-40, and incubated for 1 hour. The filters were then incubated in 1X HBB supplemented  
10 with 1% non-fat powdered milk and 0.05% NP-40 for at least 30 minutes prior to hybridization.

#### Hybridization of Filters

To prepare the hybridization solution, protein expression in E. coli (DH5a) transformed with  
15 pGEX-2TK and pGEX-2TK-derived plasmids were induced with IPTG as described above in Example 2. The bacteria were then pelleted and resuspended in 1/10 the original culture volume in Hyb 75 (20 mM Hepes [pH 7.7], 75 mM KCl, 0.1 mM EDTA, 2.5 mM MgCl<sub>2</sub>, 1 mM  
20 DTT, 0.05% NP-40). The bacterial suspension was sonicated with a probe type sonicator (Branson) and centrifuged at 10,000 x G for 5 min at 4°C. For western blots the clarified supernatant was used undiluted or diluted 1:2 with Hyb 75. For screening  
25 plaque lifts the clarified supernatant was diluted 1:2-1:8 with Hyb 75 supplemented with 1% nonfat powdered milk.

The relevant 32P-labeled GST fusion protein was added at 100,000-250,000 cpm/ml. Hybridization was  
30 carried out at 4°C with gentle rocking overnight,

-56-

after which the filters were washed three times  
(10-15 min/wash) with Hyb 75 with 1% nonfat powdered  
milk. The filters were then dried, covered with  
Saran wrap and exposed to film at -70°C with an  
05 intensifying screen.

Screening Precleared Lysates for Specifically  
Interacting Proteins

Previously, a series of GST-RB fusion proteins,  
expressed from pGEX-2T vectors, and non-covalently  
10 bound to glutathione sepharose, were used to search  
for cellular proteins capable of interacting,  
directly or indirectly, with the pRB T/E1A/E7-binding  
domain (Kaelin, W.G. *et al.*, *Cell* 64:521-532 (1991)).  
At least 7 cellular proteins were identified which  
15 bound to GST-RB fusion proteins having an intact  
T/E1A/E7-binding domain. These cellular proteins  
failed to bind to GST-RB fusion proteins (with no HMK  
sequence) derived from spontaneously occurring, loss  
of function RB mutations. In addition, the binding  
20 of these cellular proteins to GST-RB fusion proteins  
was blocked by a synthetic peptide corresponding to  
the pRB-binding/transforming sequence found in SV40 T  
antigen. In contrast, a point mutant derivative of  
this peptide, corresponding to the sequence found in  
25 the transformation defective T mutant K1, did not  
inhibit binding.

It appeared that one or more of these proteins  
may fulfill the criteria for a meaningful pRB  
cellular ligand. Unfortunately, which of the  
30 cellular proteins bound directly to the pRB T/E1A/E7-  
binding domain could not be determined from these  
experiments.

-57-

To demonstrate which of these cellular proteins is capable of interacting directly with RB, an <sup>35</sup>S-labeled WERI-Rb27 retinoblastoma cell lysate was prepared. The lysate was precleared by passage over  
05 glutathione sepharose which had been loaded with pGEX-2T-encoded GST to remove proteins which may bind to GST rather than the protein of interest. GT-RB(379-792) fusion protein was loaded onto the  
10 glutathione sepharose beads, and the precleared lysate was incubated with the bound GT-RB(379-792) fusion protein in the presence (for strip 1) or absence (for strips 2-4) of wild-type E7 peptide (residues 16-32). Unbound protein was removed by washing and the proteins were eluted by boiling in  
15 sample buffer.

Bound proteins (directly and indirectly bound to the RB-bound beads) were loaded in wide wells, resolved by SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. The filter was  
20 then cut into adjacent strips and either subjected to autoradiography (strips 1 and 2) or probed with <sup>32</sup>P-GTK-RB(379-792) in the presence (strip 3) or absence (strip 4) of the above-mentioned E7 peptide. Strips 3 and 4 were then washed, dried, placed under  
25 saran wrap (which greatly reduces the <sup>35</sup>S signal), and placed under film.

At least two bands appeared to be capable of interacting directly with the <sup>32</sup>P-GTK-RB(379-792) probe in a peptide inhibitable manner. These bands  
30 were not detected by <sup>32</sup>P-GTK-RB(379-792) on strips in which the lysate was incubated with GT-RB(379-792) in

-58-

the presence of wild-type E7 peptide prior to electrophoresis, indicating specificity of binding. Furthermore, the ability of these two cellular proteins to interact with  $^{32}\text{P}$ -GTK-RB(379-792) in the filter binding assay was not inhibited by a point mutant (Glu<sub>26</sub> to Gln<sub>26</sub>) version of the E7 peptide defective in RB binding.

Screening Whole-cell Lysates for Specifically Interacting Proteins

An unlabeled whole cell lysate prepared from WERI-rb27 cells was also probed for cellular proteins capable of binding directly to the RB. The unlabeled extract was subjected to SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose for Western blotting. The filter was cut into strips which were probed with selected  $^{32}\text{P}$ -labeled fusion proteins. Again, there appeared to be at least 2 cellular proteins capable of interacting with  $^{32}\text{P}$ -GTK-RB fusion proteins in which the T/E1A-binding region was intact ( $^{32}\text{P}$ -GTK-RB(379-792) and  $^{32}\text{P}$ -GTK-RB(379-928)pm706) fusion proteins).

Furthermore, the binding of these cellular proteins to wild-type  $^{32}\text{P}$ -GTK-RB(379-792) fusion protein could be selectively inhibited by three different RB-binding peptides. Peptide replicas of the RB-binding sequences found in T antigen (amino acids 102-115), E1A (tyrosine followed by E1A residues 115-132), or E7 (amino acids 16-32) each inhibited binding of the labeled probe to the two cellular proteins. In contrast, point mutant

-59-

derivatives of these peptides failed to inhibit the interaction of the probe with the filter-bound cellular proteins. The particular mutant peptides tested included the T antigen peptide with a Glu<sub>107</sub> to Lys change, the E1A peptide with a Cys<sub>125</sub> to Gly change, and the E7 peptide with a Glu<sub>26</sub> to Gln change. Similar results were obtained when this assay was performed with a human Burkitt's lymphoma cell line (Akata) as well as with normal peripheral blood lymphocytes.

#### Use of Labeled GST-fusion Proteins as Probes for the Isolation of Genes

In order to isolate cDNAs encoding cellular proteins capable of interacting specifically, and directly with RB, an Akata  $\lambda$ gt11 expression library was screened with the <sup>32</sup>P-GTK-RB(379-792) probe as described in the protocols above. Six  $\lambda$ gt11 clones (clones 1, 3, 4, 5, 6 and 9) encoding  $\beta$ -galactosidase fusion proteins which bound to <sup>32</sup>P-GTK-RB(379-792) with high affinity were plaque purified and subjected to further analysis.

The binding of the fusion proteins encoded by 5 of these clones (clones 1, 4, 5, 6 and 9) to the <sup>32</sup>P-GTK-RB(379-792) probe was markedly reduced or undetectable in the presence of the wild-type E7 peptide, while the mutant E7 peptide had no effect on binding. Furthermore, the fusion proteins encoded by all of the clones bound readily to <sup>32</sup>P-GTK-RB(379-928), whereas binding to <sup>32</sup>P-GTK-RB(379-928;706F), a protein with a mutation

-60-

in the T/E7/E1A binding region of RB, was undetectable. Thus, binding to these clones displayed RB binding behavior similar to cellular proteins which interact with RB.

05 Cross-hybridization experiments, and subsequent sequence analysis, demonstrated that 4 of the clones (Clones 1, 3, 4 and 6) contained overlapping cDNA fragments derived from a common mRNA. The gene encoding this mRNA will be referred to as RBAP1.

10 Sequence analysis of clone 5 predicted that it encoded a  $\beta$ -galactosidase leader polypeptide fused to the sequence His-Ser-Phe-Leu-Leu-Cys-Asp-Glu-Asn-Val-Leu-Asp-Stop. Thus, this fusion protein contains the Leu-X-Cys-X-Glu motif common to the viral RB-binding  
15 motifs. Additional cDNA clones related to clone 5 were obtained by screening a 293 cell library with the clone 5 insert cDNA. Sequence analysis of these clones suggested that the short open reading frame in clone 5 was generated by the juxtaposition of a  
20 normally untranslated cDNA segment with the  $\lambda$ gt11  $\beta$ -galactosidase coding sequence. Thus, while the clone may not represent a cellular protein, upon expression in  $\lambda$ gt11, it encodes a protein with properties (e.g., a Leu-X-Cys-X-Glu motif) common  
25 with viral RB-binding motifs.

Clone 9, unlike clone 5, contained a long open reading frame, consistent with the possibility that it encodes an RB-interacting protein.

-61-

Additional Characterization of RBAP1

The retinoblastoma gene product is believed to serve as a cell cycle regulatory element. In particular, pRB is thought to contribute to the regulation of cell cycle progression as cells traverse the G1/S boundary. Consistent with a possible role as an RB-interacting protein, RBAP1 message levels appear to respond to cell cycle events.

A Northern blot was prepared of total RNA obtained from resting peripheral blood lymphocytes (PBL) and from PBL at various time points following stimulation with a cocktail containing PMA, PHA, and a calcium ionophore. The Northern blot was probed with an RBAP1 probe from clone 4. A single message of about 3.5 kb became detectable 24-36 hours after stimulation.

In a similar experiment, PBL were stimulated to enter the cell division cycle in the presence of hydroxyurea (HU). Again, a single 3.5 kb mRNA was detected with the RBAP1 probe. The abundance of the message in RNA from HU-treated PBLs increased to a maximum at 36 hours, and then appeared to plateau. Furthermore, the level of the mRNA in cells detected by the RBAP1 probe fell dramatically within 8 hours after removal of HU.

In vitro Binding of RBAP1 to pRB

The RBAP1 cDNA was cloned into pGEX-2T and expressed in E. coli as a GST-fusion protein. Glutathione sepharose beads were loaded with the

-62-

GST-RBAP1 fusion. <sup>35</sup>S-labeled RB proteins were prepared by in vitro transcription and translation from cDNA clones as described (Kaelin, W.G. et al., Mol. Cell. Biol. 10(7): 3761-3769 (1990)). The  
05 labeled RB proteins, RB(379-928) and RB(379-928;pm706) were incubated with the bound GST-RBAP1 fusion protein. The beads were washed and analyzed for the amount of fusion protein retained. Quantitative recovery of the in vitro translated  
10 <sup>32</sup>P-GTK-RB(379-928) protein, but not of the <sup>32</sup>P-GTK-RB(379-928;pm706) fusion protein was observed. This experiment further suggests that the RBAP1 product is capable of interacting directly with the RB gene product.

15 Expression of p107 as a GTK-fusion Protein

The RB-related protein p107 was also expressed as a GTK-fusion protein. A cDNA encoding a region of p107 which is homologous to the T/E1A/E7 binding domain of the RB gene product was inserted into  
20 pGEX-2TK. Using the protocols described in the above examples, a GTK-p107 fusion protein was expressed in E. coli, a lysate was prepared, the fusion protein present in the lysate was captured on a glutathione-sepharose affinity column, and labeled  
25 with [<sup>32</sup>P] in situ in a phosphorylation reaction using the catalytic subunit of cAMP-dependent protein kinase and [ $\gamma$ -<sup>32</sup>P]ATP. The [<sup>32</sup>P]-labeled GTK-p107 fusion protein was eluted from the column, and was  
30 used as probe in Western blot analysis of whole cell extracts. Whole cell extracts (unlabeled) containing

-63-

E1A were prepared from 293 cells, subjected to SDS-polyacrylamide gel electrophoresis, and electrophoretically transferred to filters. Filters were cut into strips and probed with (1) anti-E1A antibody, (2) [<sup>32</sup>P]-labeled GTK-p107 fusion protein, (3) [<sup>32</sup>P]-labeled GTK-p107 fusion protein in the presence of competing wild-type E7 peptide (residues 16-32), and (4) [<sup>32</sup>P]-labeled GTK-p107 fusion protein in the presence of mutant E7 peptide (residues 16-32, with a Glu to Gln mutation at residue 26). Autoradiography indicated that the [<sup>32</sup>P]-labeled GTK-p107 fusion protein, alone or in the presence of mutant E7 peptide, was able to detect the E1A product on the blots. The specificity of interaction was indicated by the observation that the [<sup>32</sup>P]-labeled GTK-p107 fusion protein did not detect the E1A product in the presence of competing wild-type E7 peptide.

#### Example 5

#### 20     Construction of pAR( $\Delta$ RI)59/60 and Properties of pFEK-fusion Proteins

##### Construction of pAR( $\Delta$ RI)59/60

Plasmid pAR3040 (also referred to as pET3a) was the starting material (Studier, Meth. Enzymol .185: 60-89 (1990)). This plasmid and expression plasmids derived from this vector can be maintained and induced for expression as described by Studier (Studier, Meth. Enzymol .185: 60-89 (1990)). Plasmid

-64-

pAR3040 was cut with EcoRI. The overhangs were made blunt by "filling in" using Klenow enzyme and dNTPs, and religated to destroy the EcoRI site. In an infrequent event, the EcoRI site was regenerated on propagation of the plasmid. Therefore, another version of of this intermediate was constructed by treating with mung bean nuclease to remove the EcoRI overhangs prior to religation. Regeneration of the EcoRI site in the latter version was not observed.

Both versions of the intermediate construction were modified as described below, to make two slightly different pAR( $\Delta$ RI)59-60 vectors (i.e., differing in the manner in which the EcoRI site was destroyed). These vectors behave identically with respect to expression of encoded fusion proteins. The intermediate vectors were then cleaved with NdeI and treated with calf intestinal alkaline phosphatase (CIP).

-65-

Two complementary oligonucleotide adaptors, MAB 59 and MAB 60, were synthesized, kinased, and annealed. The adaptors were then ligated into the NdeI cleaved, CIP-treated vectors. The adaptors  
05 introduce the 17 amino acid sequence comprising the FLAG peptide and HMK site shown in Figure 2. The sequence of the adaptors is shown below:

Met Asp Tyr Lys Asp Asp Asp Asp Lys Ala Arg Arg-

5'-T ATG GAC TAC AAA GAC GAT GAC GAT AAA GCA AGA AGA-  
10 AC CTG ATG TTT CTG CTA CTG CTA TTT CGT TCT TCT-

Ala Ser Val Gln Phe-

GCA TCT GTG GAA TTC CA  
CGT AGA CAC CTT AAG GT AT-5'

Plasmids having one insert of the double-stranded  
15 adaptor were identified by restriction analysis of miniprep DNA. An XbaI-BamHI double digest releases a fragment comprising the inserted sequence. The structure of resulting plasmids, named pAR( $\Delta$ RI)59/60, was confirmed by dideoxy sequencing using  
20 oligonucleotide primers MAB 58 (5'-GCAGCCAACTCAGCTTC-3') and MAB 69 (5'-TTAATACGACTCACTAT-3').

#### Construction of Derivatives of pAR( $\Delta$ RI)59/60

Four derivatives of pAR( $\Delta$ RI)59/60 were prepared.

-66-

The first derivative encoding a selected polypeptide was prepared by the insertion of a 1.5 kb EcoRI fragment of a shPan-1 (N3) cDNA into pAR( $\Delta$ RI)59/60, which had been cleaved with EcoRI and treated with  
05 CIP. The shPan-1 (N3) cDNA encodes a DNA binding protein, as described by German et al. (German et al., Molec. Endocrinol. 5: 292-299 (1991)).

Plasmids containing the insert were detected by restriction analysis of miniprep DNA using a  
10 BamHI-StuI double digest. The correct orientation yielded a diagnostic 500 bp fragment. The sequence at the junctions between the vector and insert was verified by sequencing, using oligonucleotide primers MAB 58 and MAB 69 (see above). The plasmid is  
15 referred to herein as pFEK-shPan-1.

The second derivative of pAR( $\Delta$ RI)59/60 was made by the insertion of N3-SH, a fragment of shPan-1 generated by PCR (German et al., Molec. Endocrinol. 5: 292-299 (1991)). For cloning, the fragment was  
20 generated by PCR, using oligonucleotide primers MAB 72 (5'-GGCCGAATTCTCCTGGTCCCACGGAGACCC-3') and MAB 73 (5'-GGCCGAATTCGCCGAGGAGGACAAGAAGGACC-3'). The PCR products were purified on an agarose gel, electroeluted, treated with T4 DNA polymerase, and  
25 digested with EcoRI. The resulting fragment was ligated to EcoRI-cut, CIP-treated pAR( $\Delta$ RI)59/60. Constructs with single inserts were identified by XbaI-BamHI double digests of miniprep DNA. The junctions and sequence of the insert were verified by  
30 dideoxy sequencing using oligonucleotides MAB 58 and MAB 69 (see above). The resulting construct is referred to herein as pFEK-N3-SH.

-67-

The third derivative of pAR( $\Delta$ RI)59/60 was constructed by the insertion of a DNA sequence encoding amino acids 120 through 206 of the rat c-fos protein (Kouzarides, T. and E. Ziff, Nature 336: 646-651 (1988); Curran, T. et al., Oncogene 2: 79-84 (1987); Nakabeppu, Y. and D. Nathans, EMBO J. 8: 3833-3841 (1989)). The specific fragment used for cloning was generated by the polymerase chain reaction (PCR), using oligonucleotide primers MAB 70 (5'-GGCCGAATTCGCGCAGAGCATCGGCAGAAG-3') and MAB 71 (5'-GGCCGAATTCCTACTAGATCTTGCAGGCAGGTCGGT-3'). The resulting PCR products were purified on an agarose gel, electroeluted, treated with T4 DNA polymerase, and digested with EcoRI. The resulting fragment was ligated into EcoRI-cut, CIP-treated pAR( $\Delta$ RI)59/60. Isolates with single inserts were identified by XbaI and BamHI double digests. The junctions and insert sequence were verified by dideoxy sequencing using the oligonucleotide primers MAB 58 and MAB69 (see above). The resulting plasmid is referred to herein as pFEK-c-fos(120-206).

The fourth derivative of pAR( $\Delta$ RI)59/60 was derived by the insertion of amino acids encoding residues 206 through 340 of the human c-jun protein (Kouzarides, T. and E. Ziff, Nature 336: 646-651 (1988); Bohmann, D. et al., Science 238: 1386-1392 (1987)). The fragment for cloning was generated by PCR using oligonucleotide primers MAB 74 (5'-GGCCGAATTCCTTTCCCGCGCAACCCAGCA-3') and MAB 75 (5'-GGCCGAATTCCTTCCCGACGGTCTCTCTTCAAA-3'). The PCR products were purified on an agarose gel,

-68-

electroeluted, treated with T4 DNA polymerase, and digested with EcoRI. The resulting fragment was ligated to EcoRI-cut, CIP-treated pAR( $\Delta$ RI)59/60.

Constructs having a single insert were again  
05 identified by XbaI-BamHI digestion of mini-prep DNA. The junctions and insert sequences were confirmed by dideoxy sequencing using primers MAB 58 and MAB 69 (see above). The resulting plasmid is referred to herein as pFEK-c-jun(206-340).

10 pFEK-fusion Proteins Retain Biological Activity and Are Efficiently Phosphorylated

As each of the selected polypeptides expressed from pAR( $\Delta$ RI)59/60 is a DNA binding protein, the activity of the selected polypeptide portions, in the  
15 context of a fusion protein comprising an affinity ligand and modification site, were assayed by electrophoretic mobility shift assays (EMSA). Assay conditions for the FEK-shPAN-1 fusion protein encoded by pFEK-shPan-1 and by pFEK-N3-SH, were as described  
20 by German *et al.* (German *et al.*, Molec. Endocrinol. 5: 292-299 (1991)). EMSA conditions for the fusion protein encoded by pFEK-c-fos(120-206) were as described (Kouzarides, T. and E. Ziff, Nature 336: 646-651 (1988); Curran, T. *et al.*, Oncogene 2: 79-84  
25 (1987); Nakabeppu, Y. and D. Nathans, EMBO J. 8: 3833-3841 (1989)).

EMSA conditions for the fusion protein encoded by pFEK-c-jun(206-340) were as described (Kouzarides, T. and E. Ziff, Nature 336: 646-651 (1988); Bohmann,  
30 D. *et al.*, Science 238: 1386-1392 (1987)). The DNA

-69-

binding activity of the FEK-c-jun(206-340) fusion protein present in bacterial extracts, was assayed in combination with bacterial extracts containing fos "core" (to assay binding of jun as a jun-fos heterodimer).

The four unlabeled fusion proteins were made by in vitro transcription/translation. The particular DNA probes used in each shift assay were radiolabeled for detection of DNA-fusion protein complexes. As a positive control, each of the native (i.e., not fused) DNA binding proteins was prepared by in vitro transcription/translation and subjected to the shift assay. The short polypeptide produced by PAR( $\Delta$ RI)59/60, and the transcription/translation products from a vector control for each pFEK-construct in which the selected polypeptide insert was in a reverse orientation, provided negative controls for DNA binding. As indicated by the shift assay, all four pFEK-fusion proteins bound DNA to an extent comparable to the corresponding native protein. These results indicate that the incorporation of the selected polypeptides (fos, jun, shPan-1 (N3), and N3-SH) into fusion proteins of the present invention comprising a modification recognition site, did not alter the biological activity (DNA binding) of these proteins.

In parallel to the experiment just described, in vitro transcribed and translated fusion proteins were modified by phosphorylation with HMK kinase and non-radioactive ATP (see below for conditions). Modification by phosphorylation could be

-70-

distinguished because the negative charge of the additional phosphate group on the modified fusion protein led to an alteration in migration of the DNA-fusion protein complex as compared with the  
05 unmodified fusion protein-DNA complex. Modification of all four of the fusion proteins did not alter the extent of complex formation, indicating that the biological activity of the selected polypeptide portion is not altered by modification.

10 Induction of Protein Expression in BL21 Bacteria

Induction of expression from pAR( $\Delta$ RI)59/60 and pFEK- plasmids was done essentially as described by Studier et al. (Studier, Meth. Enzymol. 185: 60-89 (1990)). Briefly, plasmids were transformed into  
15 competent BL21 or BL21 pLySS bacteria. A single colony was inoculated into Luria broth with ampicillin at 50 ug/ml and with chloramphenicol at 25 ug/ml. Growth at 37°C was allowed to proceed until an optical density ( $A_{600}$ ) of approximately 1.0 was  
20 attained. IPTG was added as described by Studier (Meth. Enzymol. 185: 60-89 (1990)). Cells were grown for another 3 hours or so at 37°C. A bacterial lysate was prepared using standard techniques.

Following expression in E. coli, the  
25 unfractionated bacterial extract from transformants carrying expression vector pFEK-shPan-1, pFEK-N3-SH, pFEK-c-fos(120-206), or pFEK-c-jun(206-340) was assayed by electrophoretic mobility shift assay as in the previous section. The results indicated that the  
30 unmodified or modified (by phosphorylation with

-71-

unlabeled ATP) FEK-fusion proteins expressed in bacterial extracts retained the biological activity of the selected polypeptide portion.

Partial Purification of Proteins Produced:

05       Following expression in E. coli, bacterial extracts from cells transformed with a pFEK- vector were fractionated to obtain partially purified FEK-fusion proteins by one of two methods.

A)     DEAE-Sephacel chromatography:

- 10       -used as per directions of supplier  
          (Pharmacia-PL)  
      -buffer contained 100 mM NaCl and 10% glycerol  
          (in addition to the buffer constituents  
          presents in the bacterial lysates)  
15       -treated in "batch" for 60 minutes at +4°C  
          -lysate was incubated with resin with  
          continuous gentle agitation  
      -supernatant solution passed over a 2 ml  
          column of resin  
20       -flow-through collected

The fusion proteins encoded by vectors pFEK-shPan-1 and pFEK-N3-SH were obtained using DEAE-Sephacel chromatography at ~50% purity.

B)     Heparin-Sepharose chromatography:

- 25       -used as per directions of supplier  
          (Pharmacia-PL)

-72-

- DEAE flow-through fraction from protocol (A) was loaded onto a 1 ml bed volume column of resin
- resin washed with 10 column volumes of load buffer (i.e., 100 mM salt)
- proteins were eluted with 1 column volume each of 200, 300, 400, 500, 600, 700, 800, 900 and 1000 mM salt
- fractions were tested for the presence of active protein by the appropriate assay

For example, in the case of a fusion protein comprising the fos 'core' (encoded by pFEK-c-fos(120-206)), activity of fractions was assayed by electrophoretic mobility shift assay together with reticulocyte-produced c-jun protein. Peak activity for the fusion protein comprising the fos 'core' fusion protein was eluted at approximately 500-600 mM salt yielding a preparation with ~50% purity.

#### Phosphorylation of Proteins with HMK:

All four fusion proteins produced by the pFEK-vectors (pFEK-shPan-1, pFEK-N3-SH, pFEK-c-fos(120-206), and pFEK-c-jun(206-340)) were capable of being modified by phosphorylation when present in crude bacterial extracts. As discussed above, modification by phosphorylation with unlabeled ATP was observed. In addition, modification using radioactively-labeled ATP was observed.

-73-

Proteins which were partially purified by the methods (A and B) described above were also efficiently radiolabeled by phosphorylation using HMK enzyme and [ $\gamma$ - $^{32}$ P]ATP. Labeling to a specific activity of approximately  $10^7$  to  $10^8$  cpm per  $\mu$ g of protein was observed for partially purified FEK-fusion proteins encoded by pFEK-shPan-1, pFEK-N3-SH, and pFEK-c-fos(120-206). The protocols used for protein labeling are described below.

10 Phosphorylation with Heart Muscle Kinase:

Sigma # P-2645 Protein Kinase, Catalytic subunit from bovine heart was obtained as lyophilized powder in 250 unit vials. Typically, 250 units of HMK enzyme was resuspended in 25  $\mu$ l of 40 mM DTT (i.e., at 10 u/ $\mu$ l). The solution was allowed to stand at room temperature for 10 minutes and was stored at +4 °C. Activity was stable for 2-3 days, but HMK was usually freshly reconstituted for each use.

A 10x preparation of HMK Buffer was prepared (200 mM Tris-Cl (pH 7.5), 10 mM DTT, 1 M NaCl, and 120 mM MgCl<sub>2</sub>). The phosphorylation reaction mixture was as follows:

- 3  $\mu$ l (10x) HMK buffer
- 2-5  $\mu$ l of [ $\gamma$ - $^{32}$ P] (e.g., NEN-Dupont #NEG-035C [ $\gamma$ - $^{32}$ P]-ATP >7000Ci/mmol)
- 1-10  $\mu$ l of protein extract (amount will vary with the particular protein)
- 1  $\mu$ l of 10 u/ $\mu$ l HMK
- water to a total of 30  $\mu$ l

-74-

The reaction mixture was incubated at 37 °C for 30-60 minutes, and was stored on ice until use. Some proteins were labeled almost as well when the incubation was carried out on ice. The low  
05 temperature incubation can serve to preserve the biological activity of the modified fusion protein. These procedures can be scaled up as required.

G50 column chromatography to Remove Unincorporated Label

10 FEK-fusion proteins were used for Western blotting or for plaque screening. For these applications, the fusion protein was run a G50 column as described below in order to remove unincorporated label following the HMK phosphorylation reaction.

15 "Z + 0.1 M KCl" buffer (25 mM Hepes-KOH ([pH 7.7], 12.5 mM MgCl<sub>2</sub>, 20% glycerol, 100 mM KCl) was prepared. For solutions comprising Z + KCl, and BSA, the solution was filtered through a 0.2 µm filter prior to use. DTT was added to a final concentration  
20 of 1 mM just prior to use.

--G50 fine (pre-swollen in water) was washed (1X) with 5 ml of (Z + 0.1 M KCl + 3 mg/ml BSA + 1 mM DTT) and resuspended at [1:1] in the same buffer  
-the beads were rotated at RT for -1 hour, and washed  
25 three times with (Z + 0.1 M KCl + 1 mg/ml BSA + 1 mM DTT)  
-a 1 ml sterile plastic pipette was packed with the

-75-

washed G50 (bed vol -1.2 ml) and equilibrated at room temperature with -50-10 ml of (Z - 0.1 M KCl + 1 mg/ml BSA + 1 mM DTT)

- immediately prior to loading the column, the
- 05 total volume of the HMK reaction was brought to 100  $\mu$ l with ice-cold (Z + 0.1 M KCl + 1 mM DTT)
- the column was loaded and run at room temperature, taking 1 drop fractions (-45  $\mu$ l/drop)
- fractions were stored immediately on ice
- 10 -2-5  $\mu$ l aliquots were removed from each fraction and counted by Cerenkov counting
- the excluded peak fractions (usually at 1/3 to 1/2 the column volume) were pooled

- Optionally aliquots of the fractions (e.g., 1.0  $\mu$ l)
- 15 can be fractionated on SDS-polyacrylamide gels to monitor the progress of the fusion protein. The above procedure can also be carried out at 4° C where desired.

#### Equivalents

- 20 Those skilled in the art will be able to recognize, or be able to ascertain, using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be
- 25 encompassed by the following claims.

-76-

CLAIMS

1. An expression vector comprising a nucleotide  
sequence which encodes at least one affinity  
ligand, at least one modification recognition  
05 sequence and at least one restriction site for  
the in frame insertion of a nucleotide sequence  
encoding a selected polypeptide.
2. The expression vector of Claim 1 wherein the  
affinity ligand is a glutathione-S-transferase  
10 affinity ligand or a FLAG epitope affinity  
ligand, and the modification recognition  
sequence is a phosphorylation site.
3. The expression vector of Claim 1 further  
comprising a nucleotide sequence which encodes a  
15 cleavable linker located between the affinity  
ligand and the restriction site for insertion of  
a selected polypeptide.
4. The expression vector of Claim 1, wherein the  
affinity ligand is a glutathione-S-transferase  
20 affinity ligand or a FLAG epitope affinity  
ligand, the modification recognition sequence is  
a phosphorylation site, and the cleavable linker  
is selected from the group consisting of a  
thrombin cleavage site, Factor Xa cleavage site,  
25 and enterokinase cleavage site.

-77-

5. The expression vector of Claim 4, wherein the vector is pGEX-2TK.
6. The expression vector of Claim 4, wherein the vector is pAR( $\Delta$ RI)59/60.
- 05 7. An expression vector comprising a nucleotide sequence which encodes a fusion protein comprising an affinity ligand, a modification recognition sequence and a selected polypeptide.
8. The expression vector of Claim 7, wherein the  
10 affinity ligand is a glutathione-S-transferase affinity ligand or a FLAG epitope affinity ligand, and the modification recognition sequence is a phosphorylation site.
9. The expression vector of Claim 7 wherein the  
15 nucleotide sequence further encodes a cleavable linker located between the affinity ligand and the selected polypeptide.
10. The expression vector of Claim 9, wherein the  
20 affinity ligand is glutathione-S-transferase affinity ligand or a FLAG epitope affinity ligand, the modification recognition sequence is a phosphorylation site, and the cleavable linker is selected from the group consisting of a  
25 thrombin cleavage site, factor Xa cleavage site and enterokinase cleavage site.

-76-

11. A method of producing a labeled fusion protein comprising:
- 05 a) expressing a fusion protein comprising an affinity ligand, a phosphorylation recognition sequence and a polypeptide in a host cell using an expression vector;
  - b) lysing the host cells to obtain a lysate containing the fusion protein;
  - 10 c) contacting the lysate with an affinity matrix under conditions sufficient to bind the affinity ligand portion of the fusion protein to the affinity matrix;
  - d) washing the product of step (c) to remove unbound material; and
  - 15 e) contacting the bound fusion protein with [ $\gamma$ -labeled]ATP and a kinase reaction buffer comprising a protein kinase under conditions appropriate for phosphorylation of the bound fusion protein to thereby  
20 produce a labeled fusion protein which is bound to the affinity matrix.
12. The method of Claim 11 further comprising the steps of:
- 25 f) optionally stopping the reaction of step (e) by addition of a stop buffer;
  - g) washing the affinity matrix with labeled fusion protein bound thereto to remove unincorporated label; and

-79-

- h) eluting the labeled fusion protein with a suitable elution buffer comprising a release component thereby obtaining a labeled fusion protein.
- 05 13. The method of Claim 11 further comprising the steps of:
- f) optionally stopping the reaction of step (e) by addition of a stop buffer;
- 10 g) washing the bound labeled fusion protein to remove unincorporated label; and
- h) cleaving the labeled fusion protein with a site specific protease to produce a labeled selected polypeptide portion.
- 15 14. A method of producing a labeled fusion protein comprising:
- a) expressing a fusion protein comprising a GST affinity ligand, a phosphorylation recognition sequence and a polypeptide in a
- 20 host cell using an expression vector;
- b) lysing the host cells to obtain a lysate containing the fusion protein;
- c) contacting the lysate with immobilized glutathione under conditions sufficient to
- 25 bind the GST portion of the fusion protein to bind to glutathione, thereby obtaining immobilized glutathione with the bound fusion protein;

-80-

- 05           d)    washing the immobilized glutathione with  
              bound fusion protein to remove unbound  
              material; and
- e)    contacting the bound fusion protein with  
                    [ $\gamma$ -labeled]ATP and a kinase reaction buffer  
                    comprising a protein kinase under  
                    conditions appropriate for phosphorylation  
                    of the bound fusion protein to thereby  
10                produce immobilized glutathione with  
                    labeled fusion protein which is bound to  
                    immobilized glutathione.

15.   The method of Claim 14 wherein the protein  
      kinase is the catalytic subunit of a  
      cAMP-dependent protein kinase.

15   16.   The method of Claim 14 wherein the [ $\gamma$ -labeled]ATP  
      is [ $\gamma$ -<sup>32</sup>P]ATP.

17.   The method of Claim 14 further comprising the  
      steps of:

- 20           f)    optionally stopping the reaction of step  
                    (e) by addition of a stop buffer;
- g)    washing the immobilized glutathione with  
                    bound labeled fusion protein to remove  
                    unincorporated label; and
- 25            h)    eluting the bound labeled fusion  
                    protein with a solution comprising reduced  
                    glutathione thereby obtaining a labeled  
                    fusion protein.

-81-

18. The method of claim 14 further comprising the steps of:
- f) optionally stopping the reaction of step (e) by addition of a stop buffer;
  - 05 g) washing the immobilized glutathione with bound labeled fusion protein to remove unincorporated label; and
  - h) cleaving the labeled fusion protein with thrombin to produce a labeled selected
  - 10 polypeptide portion.
19. A kit for preparing a radiolabeled glutathione-S-transferase (GST) fusion protein comprising:
- a) an immobilized glutathione affinity matrix;
  - 15 b) a wash buffer which permits binding of a GST-fusion protein to the affinity matrix;
  - c) a kinase reaction buffer;
  - d) a protein kinase preparation;
  - e) an elution buffer comprising reduced
  - 20 glutathione for releasing the fusion protein from the affinity matrix; and optionally
  - f) a reaction stop buffer.
20. A labeled fusion protein comprising an affinity
- 25 ligand, a modification recognition sequence, and a polypeptide.

-82-

21. The labeled fusion protein of Claim 20, wherein  
the affinity ligand is a  
glutathione-S-transferase affinity ligand and  
the modification recognition sequence is a  
05 phosphorylation recognition sequence.
22. The labeled fusion protein of Claim 20, wherein  
the affinity ligand is the FLAG epitope, and the  
modification recognition sequence is a  
phosphorylation recognition sequence.
- 10 23. A host cell containing the expression vector of  
Claim 1.

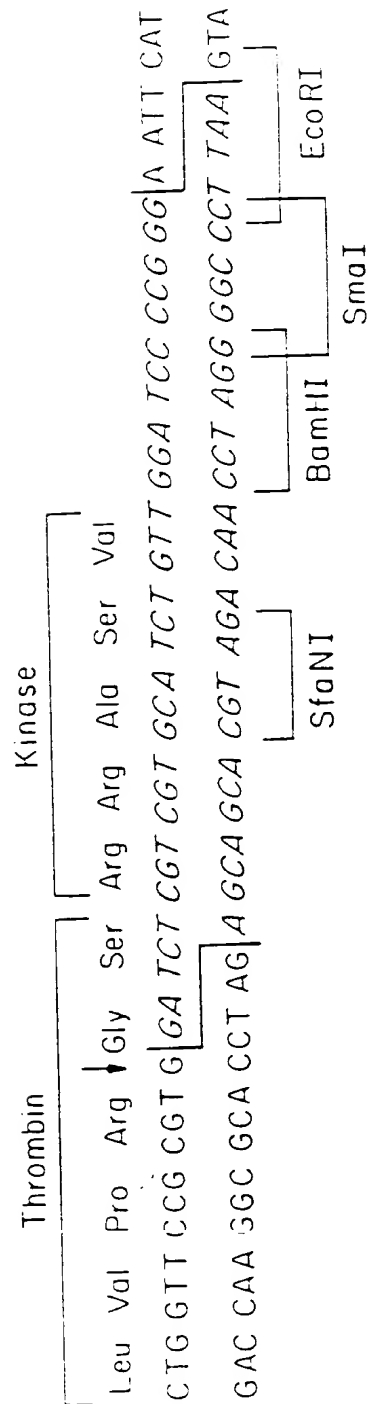


Fig. 1

2/3

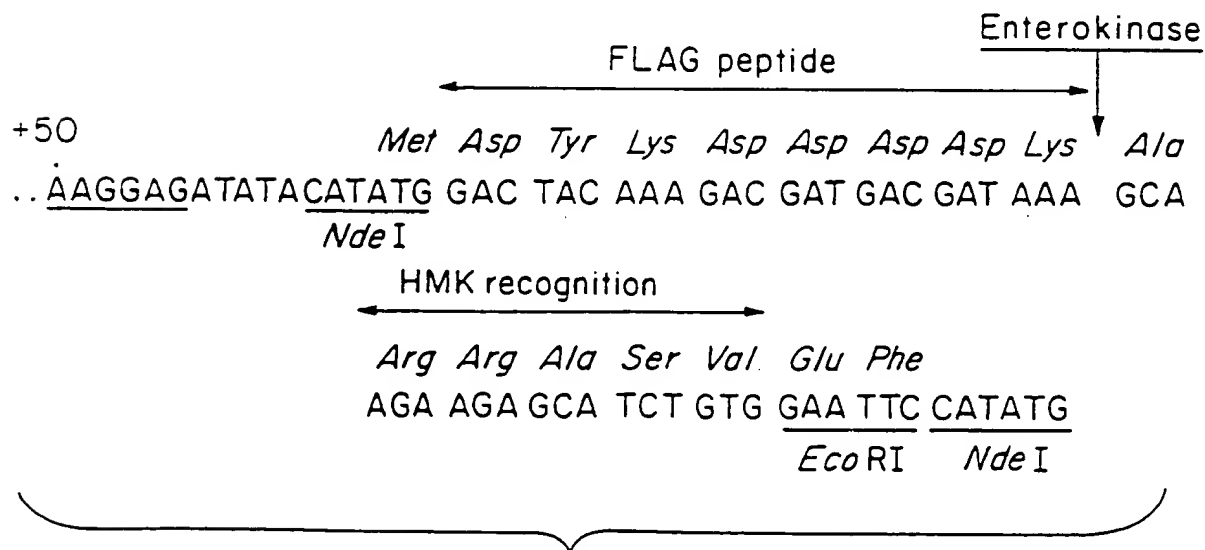
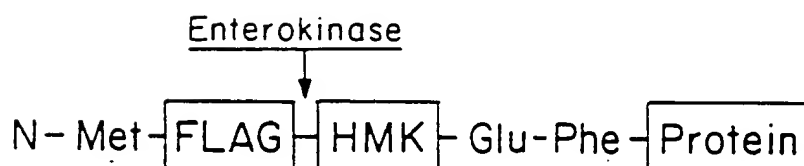
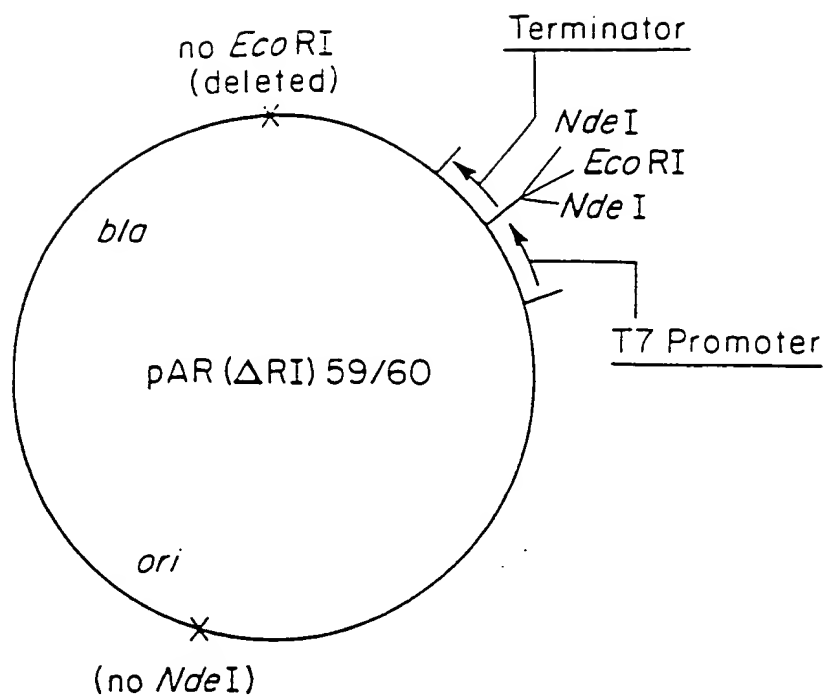
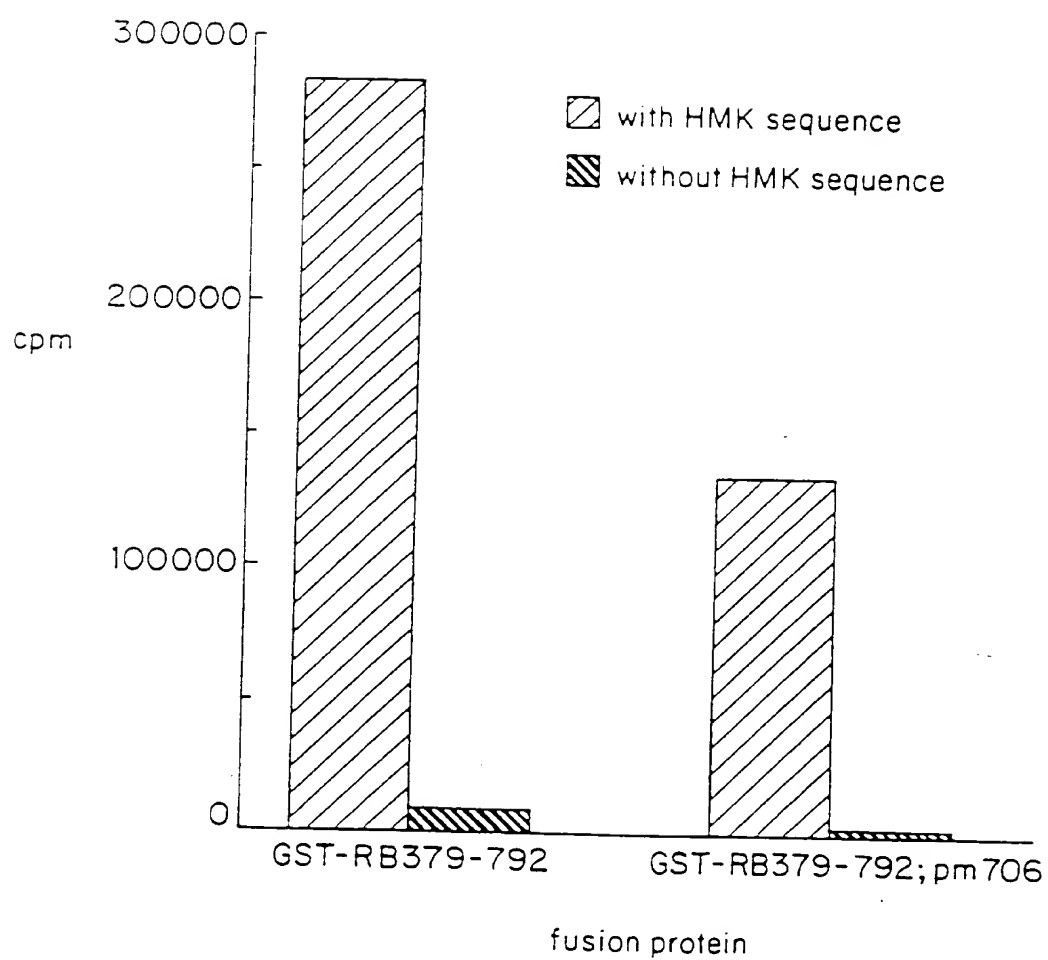


Fig. 2

*Fig. 3*

1. The first part of the document is a list of the names of the members of the committee.

2. The second part of the document is a list of the names of the members of the committee.

3. The third part of the document is a list of the names of the members of the committee.

4. The fourth part of the document is a list of the names of the members of the committee.

# INTERNATIONAL SEARCH REPORT

PCT/US 92/06187

International Application No.

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5	C12N15/62;	C12N15/12; C12N9/10; C07K13/00
II. FIELDS SEARCHED		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C07K	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched <sup>8</sup>		
III. DOCUMENTS CONSIDERED TO BE RELEVANT <sup>9</sup>		
Category <sup>10</sup>	Citation of Document <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
Y	BIO/TECHNOLOGY vol. 6, no. 10, October 1988, NATURE AMERICA, INC., NEW YORK, US pages 1204 - 1210 T.P.HOPP ET AL. 'A short polypeptide marker sequence useful for recombinant protein identification and purification' see page 1204, left column, line 1 - right column, line 47; figures 1,2 ---	1-23
Y	EP,A,0 293 249 (AMRAD CORPORATION LIMITED) 30 November 1988 cited in the application see page 3, left column, line 53 - page 5, left column, line 52; claims 1-24; figure 7 --- -/--	1-23
<p>* Special categories of cited documents: <sup>10</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"Z" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
27 OCTOBER 1992	04. 11. 92	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	HORNIG H.	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	<p>PROC. NATL. ACAD. SCI. vol. 86, no. 2, January 1989, NATL ACAD. SCI., WASHINGTON, DC, US; pages 558 - 562 B.-L. LI ET AL. 'Creation of phosphorylation sites in proteins: Construction of a phosphorylatable human interferon alpha' cited in the application see page 558, left column, line 38 - page 560, left column, line 31; figures 1,2 ---</p>	1-23
Y	<p>BIOCHEMISTRY -- vol. 29, no. 39, 2 October 1990, AM. CHEM. SOC., EASTON, PA, US; pages 9274 - 9281 N.P. GERARD AND C. GERARD 'Construction and expression of a novel recombinant anaphylatoxin, c5a-N19, as a probe for the human C5a receptor' see page 9275, left column, line 29 - line 56; figure 1 -----</p>	1-23

US 9206187  
SA 62989

.27/10/92

EP-A-0293249	30-11-88	AU-B-	607511	07-03-91
		AU-A-	1793288	21-12-88
		WO-A-	8809372	01-12-88
		JP-T-	1503441	22-11-89

